

### Protein Sequence Searches - 10/8/04

All of the sequence databases on Compugen/ABSS have been updated. A change has occurred in the protein databases.

- Two protein databases, SPTREMBL and SwissProt, are now produced as a single, merged database called UniProt.
- Results from UniProt have the file extension **.rup**.
- Sequences in UniProt are identified by the same ID that had been used in SPTREMBL or SwissProt.

If you have any questions regarding these changes or your results, please contact any STIC searcher.

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 120.835 Seconds  
(without alignments)  
5233.063 Million cell updates/sec

Title: SEQ1-A  
Perfect score: 5835  
Sequence: 1 MAWKILPIYLLLLSVFIQ.....DMDYLPVNPQGIINPMLS 1099

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 02:.\*  
1: uniprot\_sprot:.\*  
2: uniprot\_trembl:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5820.9	99.8	1404	2	Q92954	Q92954 homo sapien
2	5797.9	99.4	1404	2	Q9BX49	Q9BX49 homo sapien
3	3417	58.6	933	2	Q6ZM25	Q6ZM25 homo sapien
4	3417	58.6	933	2	BAD18580	BAD18580 h cdna fl
5	2622	44.9	1054	2	Q9JMS9	Q9JMS9 mus musculu
6	1355.4	23.2	1225	2	Q9VRA9	Q9VRA9 drosophila
7	1354.4	23.2	5173	1	MUC2_HUMAN	Q02817 homo sapien
8	1337.7	22.9	1761	2	Q7KTF6	Q7KTF6 drosophila
9	1337.7	22.9	1761	2	AAS64673	AAS64673 drosophil
10	1286	22.0	3150	2	Q7PMD5	Q7PMD5 anopheles g
11	1185.8	20.3	3409	2	Q6SSB6	Q6SSB6 chlamydomon
12	1185.8	20.3	3409	2	AAS07044	AAS07044 chlamydom
13	1143.3	19.6	1664	1	SLP1_CLOTH	Q06852 clostidium
14	1117.1	19.1	1349	2	Q8WWQ4	Q8WWQ4 homo sapien
15	1112.3	19.1	3432	2	Q9IR51	Q9IR51 drosophila
16	1112.3	19.1	3458	2	Q9IR52	Q9IR52 drosophila
17	1110.2	19.0	3889	2	Q6SSS8	Q6SSS8 chlamydomon
18	1110.2	19.0	3889	2	AAS07042	AAS07042 chlamydom
19	1087.6	18.6	1795	2	Q76894	Q76894 drosophila
20	1072	18.4	9234	2	Q7KTF5	Q7KTF5 drosophila
21	1072	18.4	9234	2	AAN10531	AAN10531 drosophil
22	1063.3	18.2	1079	2	Q9N4S7	Q9N4S7 caenorhabdi
23	1034.5	17.7	2284	2	Q9VPG1	Q9VPG1 drosophila
24	1027.6	17.6	23015	2	Q8IQ18	Q8IQ18 drosophila
25	1027.6	17.6	23015	2	AAN10358	AAN10358 drosophil
26	1024	17.5	1607	2	Q8H6QS	Q8H6QS phytoptior
27	1024	17.5	1607	2	AAP74661	AAP74661 phytoptior
28	1020.5	17.5	972	2	Q7QKK7	Q7QKK7 anopheles g
29	1015.5	17.4	1274	2	Q20007	Q20007 caenorhabdi
30	1014.9	17.4	2187	2	P70670	P70670 mus musculu
31	1014.7	17.4	10578	2	Q8ISF5	Q8ISF5 caenorhabdi

## RESULT 1

ID	Q92954	PRELIMINARY;	PRT;	1404 AA.
AC	Q92954;			
DT	01-FEB-1997 (TREMELrel. 02, Created)			
DT	01-FEB-1997 (TREMELrel. 02, Last sequence update)			
DT	01-MAR-2004 (TREMELrel. 26, Last annotation update)			
DE	Megakaryocyte stimulating factor.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Keller K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;			
RT	"Purification, Biochemical Characterization, and Cloning of a Novel			
RT	Megakaryocyte Stimulating Factor that has Megakaryocyte Colony			
RT	Stimulating Activity.;"			
RL	Blood 78:279-279(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,			
RA	Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,			
RA	Jacobs K., Turner K.;			
RT	"A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.;"			
RL	(In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,			
RL	Mosher D.F. (eds.);			
RL	BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier			
RL	Science Publishers B.V. (1993).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,			
RA	Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,			
RA	Ferez C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,			
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;			
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; U70136; AAB09089.1; -.			
DR	HSSP; P04004; 10C0.			
DR	Genew; HGNC:9364; PRG4.			
DR	GO; GO:0008283; P:cell proliferation; TAS.			
DR	InterPro; IPR000585; Hemopexin.			
DR	InterPro; IPR001212; Somatomedin_B.			
DR	Pfam; PF00045; Hemopexin; 2.			
DR	Pfam; PF01033; Somatomedin B; 2.			
DR	PRINTS; PR00022; SOMATOMEDINB.			
DR	SMART; SM00120; HX; 2.			
DR	SMART; SM00201; SO; 2.			
DR	PROSITE; PS00024; SOMATEXIN; UNKNOWN 1.			
DR	PROSITE; PS00524; SOMATEXIN B; 2.			
DR	SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;			
DR	SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;			

32	1014.7	17.4	18519	2	Q8ISF6	Q8ISF6 caenorhabdi
33	1014.7	17.4	18534	2	Q8ISF7	Q8ISF7 caenorhabdi
34	1013.8	17.4	34350	2	Q8WZ42	Q8WZ42 homo sapien
35	1009.9	17.3	1489	2	Q96449	Q96449 phytoptior
36	1008.4	17.3	7962	2	Q10465	Q10465 homo sapien
37	998.1	17.1	5703	1	MUSB_HUMAN	Q9HC84 homo sapien
38	996.1	17.1	2112	2	Q9VEL9	Q9VEL9 drosophila
39	991.2	17.0	1480	2	Q9LIE8	Q9LIE8 arabisdopsis
40	982.1	16.8	2042	2	Q767L8	Q767L8 sus scrofa
41	982.1	16.8	2042	2	BAD08434	Bad08434 sus scrof
42	967.7	16.6	926	2	Q9VVG2	Q9VVG2 drosophila
43	966.6	16.6	1121	2	Q7Z884	Q7Z884 candida alb
44	965.6	16.5	3166	2	Q9W3Z0	Q9W3Z0 drosophila
45	963.4	16.5	1720	2	Q81486	Q81486 plasmodium

## ALIGNMENTS

Query Match 99.8%; Score 5820.9; DB 2; Length 1404;  
Best Local Similarity 96.4%; Pred. No. 4.2e-103;  
Matches 1099; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQ----- 25  
DB 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60

QY 26 -----ELSCKGRCFESFERGREGCDCAOCKYDKCCPDYEFSCAEVHNPTSPSSKKAP 79  
DB 61 KRVCCTAELSCKGRCFESFERGREGCDCAOCKYDKCCPDYEFSCAEVHNPTSPSSKKAP 120

QY 80 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHSVSENOESSSSSSSSSSSTIW 139  
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHSVSENOESSSSSSSSSSSTIW 180

QY 140 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDAGSLDNGDFKVTTPDTST 199  
DB 181 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDAGSLDNGDFKVTTPDTST 240

QY 200 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTITNKQSTDG 259  
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTITNKQSTDG 300

QY 260 KEKTTSAKETSQISAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 319  
DB 301 KEKTTSAKETSQISAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 360

QY 320 TTPKEPTTTPKSAPTTPKEPATTTKSAPTTPKEPATTTKBPATTTKBPATTTKBP 379  
DB 361 TTPKEPTTTPKSAPTTPKEPATTTKSAPTTPKEPATTTKBPATTTKBPATTTKBP 420

QY 380 APTTTSKAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 439  
DB 421 APTTTSKAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 480

QY 440 EPAPTAPKAPATTPKEPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPAT 499  
DB 481 EPAPTAPKAPATTPKEPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPAT 540

QY 500 TTKSAPTTPKEPSPTTKBPATTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 559  
DB 541 TTKSAPTTPKEPSPTTKBPATTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 600

QY 560 APTAPKEPATTPKETAPTTPKKLTPTTPEKLAPTTPKEPATTTPELAATTPKEEPT 619  
DB 601 APTAPKEPATTPKETAPTTPKKLTPTTPEKLAPTTPKEPATTTPELAATTPKEEPT 660

QY 620 PREPATTPKAAAPNTPKBPATTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 679  
DB 661 PREPATTPKAAAPNTPKBPATTPKEPATTTKBPATTTKBPATTTKBPATTTKBP 720

QY 680 APTTPKAPKELAPATTTKEPTSTSDKAPATTPKGTAPTTPKEPATTTKBPATTPKG 739  
DB 721 APTTPKAPKELAPATTTKEPTSTSDKAPATTPKGTAPTTPKEPATTTKBPATTPKG 780

QY 740 TAPTTLKEPATTPKBPAPKELAPATTTKGTPTSTSDKAPATTPKETAPTTPKEPATTTK 799  
DB 781 TAPTTLKEPATTPKBPAPKELAPATTTKGTPTSTSDKAPATTPKETAPTTPKEPATTTK 840

QY 800 KBAPTTPPEPTTTSVSTPTTTKEPTTIHKSPDESTPELSABPTPKALENSKPECVPT 859  
DB 841 KBAPTTPPEPTTTSVSTPTTTKEPTTIHKSPDESTPELSABPTPKALENSKPECVPT 900

QY 860 TKTPAATKEMTTAKDKTTERDLRTTPETTTAAAPKMTKETATTTKTESKITATTTQV 919  
DB 901 TKTPAATKEMTTAKDKTTERDLRTTPETTTAAAPKMTKETATTTKTESKITATTTQV 960

QY 920 TSTTTQDTTPFKITLLTLLAPKVTITTKTITTTTTEIMNKPBEATPKDRATNSKATTPK 979  
DB 961 TSTTTQDTTPFKITLLTLLAPKVTITTKTITTTTTEIMNKPBEATPKDRATNSKATTPK 1020

QY 980 PQKPTKAPKPTSTKPKTWPVRVKPTTTPRKMSTMPELNPTSRIAEAMLQTTTRPN 1039

Db 1021 FOKPTKAPKPTSTTKKPTMPRVRKPTTTPRKMSTMPELNPTSRIAEAMLQTTTRN 1080

QY 1040 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1099

Db 1081 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1140

RESULT 2

Q9BX49 PRELIMINARY; PRT; 1404 AA.

AC Q9BX49;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

GN BG17416.2 (MSF: megakaryocyte stimulating factor ).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Wray P.;

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL133553; CAC36090.1; -.

DR HSP; P04004; 1000.

DR InterPro; IPR000585; Hemopexin.

DR InterPro; IPR001212; Somatomedin\_B.

DR Pfam; PF00045; Hemopexin; 2.

DR Pfam; PF01033; Somatomedin\_B; 2.

DR PRINTS; PR00022; SOMATOMEDINB.

DR SMART; SM00120; HX; 2.

DR SMART; SM00201; SO; 2.

DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.

DR PROSITE; PS00524; SOMATOMEDIN B; 2.

SQ SEQUENCE 1404 AA; 151076 MW; 782A11746B3FDEES CRC64;

Query Match 99.4%; Score 5797.9; DB 2; Length 1404;  
Best Local Similarity 96.1%; Pred. No. 1.1e-102;  
Matches 1096; Conservative 0; Mismatches 3; Indels 41; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQ----- 25  
DB 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60

QY 26 -----ELSCKGRCFESFERGREGCDCAOCKYDKCCPDYEFSCAEVHNPTSPSSKKAP 79  
DB 61 KRVCCTAELSCKGRCFESFERGREGCDCAOCKYDKCCPDYEFSCAEVHNPTSPSSKKAP 120

QY 80 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHSVSENOESSSSSSSSSSSTIW 139  
DB 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEITEHSVSENOESSSSSSSSSSSTIW 180

QY 140 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDAGSLDNGDFKVTTPDTST 199  
DB 181 KIKSSKNSAANRELQKLVKONKNRTKKPTPKPPVVDAGSLDNGDFKVTTPDTST 240

QY 200 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTITNKQSTDG 259  
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVKNKETTIVETKETTITNKQSTDG 300

QY 260 KEKTTSAKETSQISAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 319  
DB 301 KEKTTSAKETSQISAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPAS 360

QY 320 TTPKEPTTTPKSAPTTPKEPATTTKSAPTTPKEPATTTKBPATTTKBPATTTKBP 379  
DB 361 TTPKEPTTTPKSAPTTPKEPATTTKSAPTTPKEPATTTKBPATTTKBPATTTKBP 420

QY 380 APTTTSKAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 439  
DB 421 APTTTSKAPTTPKEPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTK 480





```

||||| 471 ATTTEKTESKITATTTQVSTTTQDTTPFKITLTKITLAPKVTYKTKITTEIMNKP 530
||||| 961 EETAKPKDRATNSKATTPKQKPTKAPKKTSTTKKPTMPRVKXPTTTPRKMTSTMP 1020
||||| 531 EETAKPKDRATNSKATTPKQKPTKAPKKTSTTKKPTMPRVKXPTTTPRKMTSTMP 590
||||| 1021 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 1080
||||| 591 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 650
||||| 1081 MDYLRVNPQGIINPMLS 1099
||||| 651 MDYLRVNPQGIINPMLS 669

RESULT 4
BAD18580
ID BAD18580 PRELIMINARY; PRT; 933 AA.
AC BAD18580;
DT 12-MAY-2004 (TrEMBLrel. 27, Created)
DT 12-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DT 12-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE CDNA FLJ16561 fis, clone SYN04003981, moderately similar to Homo
DE sapiens proteoglycan 4, (megakaryocyte stimulating factor, articular
DE superficial zone protein, camptodactyly, arthropathy, coxa vara,
DE pericarditis syndrome) (PRG4).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kituchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RA "NEDO human cDNA sequencing project.";
RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131434; BAD18580.1; -.
SQ SEQUENCE 933 AA; 102512 MW; B803773C5BC21A23 CRC64;

Query Match 58.6%; Score 3417; DB 2; Length 933;
Best Local Similarity 60.8%; Pred. NO. 1.2e-57;
Matches 668; Conservative 0; Mismatches 1; Indels 430; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQBLSCGRCFESFERGECDDAQCKYDKCCPDYE 60
DB 1 MAWKTLPIYLLLLSVFVIQVSSQBLSCGRCFESFERGECDDAQCKYDKCCPDYE 60

QY 61 SFCAEVHNTPSPSKAPPSPGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVS 120
DB 61 SFCAEVHNTPSPSKAPPSPGASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVS 120

QY 121 ENQESSSSSSSSSTTIKIKSSKNANRELOKLVKDNKNTKKKTPKPPVDE 180
DB 121 ENQESSSSSSSSSTTIKIKSSKNANRELOKLVKDNKNTKKKTPKPPVDE 180

QY 181 AGSGLDNGDFKVTTPDSTTQHKVSTSPKITTAKPINRPSLPNPSDTSKTSITVKN 240
DB 181 AGSGLDNGDFKVTTPDSTTQHKVSTSPKITTAKPINRPSLPNPSDTSKTSITVKN 240

QY 241 TVVETKETTNNQSTGCKNTTSKETSQTSKTSKTSKTSKTSKTSKTSKTSKTSK 300
DB 241 TVVETKETTNNQSTGCKNTTSKETSQTSKTSKTSKTSKTSKTSKTSKTSKTSK 300

QY 301 PALTTPKEPTTPPKPEASTTKEPTTIKSAPTTPKEPATTPKSAPTTPKEPATTT 360
DB 301 PALTTPKEPTTPPKPEASTTKEPTTIKSAPTTPKEPATTPKSAPTTPKEPATTT 360

```

```

QY 361 KEPAPTTPKGPAPTTTKEPATTTTTSAPTTPKEPATTPPKAPATTPKEPATTPKEPTP 420
DB 361 KEPAPTTPKE----- 370
QY 421 TTPKEPATTPKEPATTPKGPAPTAPKAPATTPKGPATTPKGPATTPKGPATTPKEPTPKE 480
DB 371 ----- 370
QY 481 PAPTTTKSAPTTPKGPATTTKSAPTTPKGPSTTTTKEPATTPKEPATTPPKKAPATTP 540
DB 371 ----- 370
QY 541 KEPAPTTPKEPATTTTCKKAPATAPKGPATTPKAAANTPKGPATTPKGPATTPKEPATTP 600
DB 371 ----- 370
QY 601 PTTPEELAPTTPEEPTTTPKEPATTPKAAANTPKGPATTPKGPATTPKGPATTPKEPATTP 660
DB 371 ----- 370
QY 661 KETAPTTPKGTAPTTLKEPATTPPKKAPKELAPTTTKEPTSTTSKAPATTPKGTAPT 720
DB 371 ----- 370
QY 721 PKGPATTPKEPATTPKGTAPTTLKEPATTPPKKAPKELAPTTTKEPTSTTSKAPATTP 780
DB 371 ----- 370
QY 781 TPKETAPTTPKGPATTPPKKAPATTPETPPTTSEVSTPTTKEPTTIHKSPDESTPELS 840
DB 371 -----PAPTTPPTTSEVSTPTTKEPTTIHKSPDESTPELS 410
QY 841 AEPTPKALENSKPGVPTTKTAPATKPEMTTAKDTERDLRTTPETTAAAPKMTKET 900
DB 411 AEPTPKALENSKPGVPTTKTAPATKPEMTTAKDTERDLRTTPETTAAAPKMTKET 470
QY 901 ATTTEKTESKITATTTQVSTTTQDTTPFKITLTKITLAPKVTYKTKITTEIMNKP 960
DB 471 ATTTEKTESKITATTTQVSTTTQDTTPFKITLTKITLAPKVTYKTKITTEIMNKP 530
QY 961 EETAKPKDRATNSKATTPKQKPTKAPKKTSTTKKPTMPRVKXPTTTPRKMTSTMP 1020
DB 531 EETAKPKDRATNSKATTPKQKPTKAPKKTSTTKKPTMPRVKXPTTTPRKMTSTMP 590
QY 1021 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 1080
DB 591 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 650
QY 1081 MDYLRVNPQGIINPMLS 1099
DB 651 MDYLRVNPQGIINPMLS 669

RESULT 5
QY 90999 PRELIMINARY; PRT; 1054 AA.
AC QY90999;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-NAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus.
GN Name=Prg4;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20573856; PubMed=11124536;
RA Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT "Isolation, characterization and mapping of the mouse and human PRG4
RT (proteoglycan 4) genes.";

```

```

RL Cytogenet. Cell Genet. 90:291-297(2000).
DR EMEL; AB034730; BAA92310.1; -.
DR HSSP; P04004; 10C0.
DR MGD; MGI:1891344; PrG4.
DE InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemopexin; 2.
DE Pfam; PF01033; Somatomedin_B; 2.
DE PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 2.
DR PROSITE; PS00524; SOMATOMEDINB.
SQ SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42282335 CRC64;

Query Match      44.9%; Score 2622; DB 2; Length 1054;
Best Local Similarity 46.7%; Pred. No. 1.8e-42;
Matches 541; Conservative 41; Mismatches 147; Indels 430; Gaps 23;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25
DB 1 MGWKILPVCLSLLPVLIQVSSQDLSSCAGRCGEGYSRDATCNQYHCQHYMECCPDF 60
QY 26 -----ELSCGRCFESPERGECDCDAOCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79
DB 61 KRVCSPELSCGRCFESFARGRECDQSQCKQKCCADYDFCEVHNSTSPSKTRAP 119
QY 80 PPSGASQTIKSTTKSPKPPNNKKTKVIESBEITEHSVSENQSSSSSSSSSSSTTW 139
DB 120 TPAGASDIKSTTKSPKSPST-TRTIKVVESELTTEHSSENQSSSSSSSSSS--TIR 176
QY 140 KIKSKNSANRELQKLVKONKQNRKKTKTPKPPVVDVDEAGSLDNGDFKVT--TPPT 197
DB 177 KIKSKNS-ANRELQKNPNVQKNTPKPKKNPEPPAVDEAGSLDNGEFKLTPPPDP 235
QY 198 STQNKVSTSPKITTAKPINRPSLPNPSDKETSLVWNETTVETKETTNNKQTS 257
DB 236 PTPHSKVAISPTTAAKVPVTPKPSLAPNSETSKASLASNETTVETKETTATNKQSSA 295
QY 258 DGKEKTSKAKTSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTPKPEP 317
DB 296 -SKKKTTSVKETSAEKTSKDVETS----- 321
QY 318 ASTTPKEPTTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKPEAPTT-----PK 369
DB 322 -----TPKNSAPTTTKPEVTTTKESKFLPLPQ 349
QY 370 EPAPTTTKBPAPTTKSAPTTKPEAPTTPKKAPTTTPKPEAPTTTPKEPTTPKPEAPT 429
DB 350 EPBPTAKGPPPTTKKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTK 409
QY 430 TKPEAPTTKPEAPTTAPKAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKPEAPTTKSA 489
DB 410 TKKPEPTTKPEGPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTK 461
QY 490 PTTTKPEAPTTKSAPTTKPEPTTKPEAPTTTPKPEAPTTTPKAPTTPKKAPTTPKPEAPTTPK 549
DB 462 -----EPTTPKEPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTK 507
QY 550 EPAPTTTKKAPTAAPKPEAPTTTPKETAPTTPKLTPPTPEKLAPTTPEKAPTTPELAP 609
DB 508 EPBPT----- 512
QY 610 TTPEEPTTPTPERAPTTPKAAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTPKSTAPTTPK 669
DB 513 -----TPKEBPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTKPEPTTK 538
QY 670 GTAPTTLKPEAPTTPKKAPKELAPTTTKETSTTSKAPTTTPKGTAPTTTPKPEAPTTTP 729
DB 539 -----KEPEPTTPKPK-----EPTTPKEPVPTTP 562
QY 730 KEAPTTTPKGTAPTTLKPEAPTTPKKAPKELAPTTTKGTSTTSKAPTTTPKPEAPTTTP 789

```

## RESULT 6

Q9VR49

ID Q9VR49 PRELIMINARY; PRT; 1225 AA.

AC Q9VR49;

DT 01-WAY-2000 (TREMBLrel. 13, Created)

DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)

DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)

DE CG3047-PA.

GN Name=Sgsl; ORFNames=CG3047;

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.P., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Markulov G., Malsbina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., M.G.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,



QV 1073 FMPEVTPDMXYLPRVPNQ 1090  
 Db 1150 ---PVVPQWPMQPGYPNQ 1164

## RESULT 7

MUC2\_HUMAN  
 ID MUC2\_HUMAN STANDARD; PRT; 5179 AA.  
 AC Q02817; Q14878;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Mucin 2 precursor (Intestinal mucin 2).  
 GN Name=MUC2; Synonyms=SMUC;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Intestine;  
 RX MEDLINE=94132002; PubMed=8300571;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;  
 RT "Molecular cloning of human intestinal mucin (MUC2) cDNA."  
 RI Identification of the amino terminus and overall sequence similarity  
 RI to prepro-von Willebrand factor.";  
 RL J. Biol. Chem. 269:2440-2446(1994).  
 [2]  
 RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=93016075; PubMed=1400449;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,  
 RA Kim Y.S.;  
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located  
 RT both upstream and downstream of its central repetitive region.";  
 RL J. Biol. Chem. 267:21375-21383(1992).  
 [3]  
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.  
 RX MEDLINE=91358717; PubMed=1885763;  
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,  
 RA Petersen G.M., Kim Y.S.;  
 RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays  
 RT and polymorphism.";  
 RL J. Clin. Invest. 88:1005-1013(1991).  
 CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and  
 CC other mucus membrane-containing organs. Thought to provide a  
 CC protective, lubricating barrier against particles and infectious  
 CC agents at mucosal surfaces.  
 CC -!- SUBUNIT: Multimeric.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors,  
 CC bronchus, cervix and gall bladder.  
 CC -!- PTM: All cysteine residues are involved in intrachain or  
 CC interchain disulfide bonds (By similarity).  
 CC -!- POLYMORPHISM: The number of repeats is highly polymorphic and  
 CC varies among different alleles.  
 CC -!- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
 CC -!- SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.  
 CC -!- SIMILARITY: Contains 2 WVFC domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L21998; AAB95295.1; -;  
 CC EMBL; M74027; AAA59875.1; -;  
 CC EMBL; M94131; AAA59163.1; -;  
 CC EMBL; M94132; AAA59164.1; -;

PIR: A49963; A43932.  
 Genew; HGNC:7512; MUC2.  
 MIM; 158370;  
 DR InterPro; IPR002919; Cysrich\_TIL.  
 DR InterPro; IPR006208; Cys\_knot.  
 DR InterPro; IPR006207; Cys\_knot\_C.  
 DR InterPro; IPR006209; EGF\_like-  
 DR InterPro; IPR001007; VWF\_C.  
 DR InterPro; IPR001846; VWF\_D.  
 DR Pfam; PF00007; Cys\_knot; 1.  
 DR Pfam; PF01826; TIL; 1.  
 DR Pfam; PF00093; VWC; 1.  
 DR Pfam; PF00094; VWD; 4.  
 DR SMART; SM00041; CT; 1.  
 DR SMART; SM00214; VWC; 2.  
 DR SMART; SM00216; VWD; 4.  
 DR PROSITE; PS01185; CTCK\_1; 1.  
 DR PROSITE; PS01225; CTCK\_2; 1.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 DR PROSITE; PS01308; VWF\_1; 2.  
 DR PROSITE; PS0184; VWF\_2; 2.  
 KW Glycoprotein; Repeat; Signal.  
 FT SIGNAL 1 20 Potential.  
 FT CHAIN 21 5179 Mucin 2.  
 FT DOMAIN 1401 1747 APPROXIMATE REPEATS.  
 FT REPEAT 1401 1416 1.  
 FT REPEAT 1417 1432 2.  
 FT REPEAT 1433 1448 3.  
 FT REPEAT 1449 1464 4.  
 FT REPEAT 1465 1471 5.  
 FT REPEAT 1472 1478 6.  
 FT REPEAT 1479 1494 7A.  
 FT REPEAT 1495 1517 7B.  
 FT REPEAT 1518 1533 8A.  
 FT REPEAT 1534 1556 8B.  
 FT REPEAT 1557 1572 9A.  
 FT REPEAT 1573 1596 9B.  
 FT REPEAT 1597 1612 10A.  
 FT REPEAT 1613 1635 10B.  
 FT REPEAT 1636 1651 11A.  
 FT REPEAT 1652 1675 11B.  
 FT REPEAT 1676 1683 12.  
 FT REPEAT 1684 1699 13.  
 FT REPEAT 1700 1715 14.  
 FT REPEAT 1716 1731 15.  
 FT REPEAT 1732 1747 16.  
 FT DOMAIN 4815 4886 WVFC 1.  
 FT DOMAIN 4924 4991 WVFC 2.  
 FT DOMAIN 5075 5160 CTCK.  
 FT DISULFID 5075 5122 By similarity.  
 FT DISULFID 5089 5136 By similarity.  
 FT DISULFID 5098 5152 By similarity.  
 FT DISULFID 5102 5154 By similarity.  
 FT DISULFID ? 5159 By similarity.  
 FT CARBOHYD 163 163 N-linked (GlcNAc...)  
 FT CARBOHYD 423 423 N-linked (GlcNAc...)  
 FT CARBOHYD 670 670 N-linked (GlcNAc...)  
 FT CARBOHYD 770 770 N-linked (GlcNAc...)  
 FT CARBOHYD 894 894 N-linked (GlcNAc...)  
 FT CARBOHYD 1139 1139 N-linked (GlcNAc...)  
 FT CARBOHYD 1154 1154 N-linked (GlcNAc...)  
 FT CARBOHYD 1215 1215 N-linked (GlcNAc...)  
 FT CARBOHYD 1230 1230 N-linked (GlcNAc...)  
 FT CARBOHYD 1246 1246 N-linked (GlcNAc...)  
 FT CARBOHYD 1787 1787 N-linked (GlcNAc...)  
 FT CARBOHYD 1820 1820 N-linked (GlcNAc...)  
 FT CARBOHYD 4339 4339 N-linked (GlcNAc...)  
 FT CARBOHYD 4351 4351 N-linked (GlcNAc...)  
 FT CARBOHYD 4362 4362 N-linked (GlcNAc...)  
 FT CARBOHYD 4373 4373 N-linked (GlcNAc...)  
 FT CARBOHYD 4422 4422 N-linked (GlcNAc...)  
 FT CARBOHYD 4438 4438 N-linked (GlcNAc...)  
 FT CARBOHYD 4502 4502 N-linked (GlcNAc...)

```

FT CARBOHYD 4616 4616 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 4627 4627 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 4752 4752 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 4787 4787 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 4881 4881 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 4888 4888 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 4955 4955 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 4970 4970 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 5019 5019 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 5038 5038 N-linked (GlcNAc... ) (Potential).
FT CARBOHYD 5069 5069 N-linked (GlcNAc... ) (Potential).
FT CONFLICT 1351 1351 H -> L (in Ref. 3).
FT CONFLICT 1412 1412 T -> S (in Ref. 3).
FT CONFLICT 1449 1449 L -> P (in Ref. 3).
FT CONFLICT 1504 1504 M -> T (in Ref. 3).
FT CONFLICT 4192 4192 G -> S (in Ref. 2).
SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;

Query Match
Best Local Similarity 23.2%; Score 1354.4; DB 1; Length 5179;
Matches 431; Conservative 70; Mismatches 478; Indels 796; Gaps 56;

QY 2 AKWLPY-----LULLSVFV-----QQVSSQE 26
Db 1020 SKEAPTCDVSTNPEPCSLNPHRRSWAERQCSILKSSVFSICHSKVDPKPFYEACVHDS 1079

QY 27 LSKC--GRCFESFERGECDDA-----OCKYDKC-----CPDYSEFCAEVHNPT 70
Db 1080 CSDTGGDC-----ECFSAVASAQCTKEGACVFWRTPDLCF-----IFC-DYNNPP 1127

QY 71 SP-----PSSKKAPP-----80
Db 1128 HECWEHYPCGNRSFETCRTINGIHSNISVSYLEGCVPRCPKDRPIYEBDLKKCVTADCK 1187

QY 81 -----PSGAS-----QTIKS-----90
Db 1188 GCVEDTHVPPGASVPEETCKSCVCTNSQVQVCRPEEGKILNQTDGAFYWEICGPNG 1247

QY 91 -----TTKRSPKPNKKTKKVIBSEBITEHSVSENQESSSS 129
Db 1248 TVEKHNI CSITRSTLTFTTITLPTTSTT-----LSTPKLCLLWSDMINEHPS 1315

QY 130 SSSSSSTIWKIKSNAANRELQKLKVDKNKRNKTKKTKPKPPV-----DEAG 182
Db 1286 TITPTSSSTV-----VTPTDSTTQH-NKVSISPK-----210

QY 183 SGLDNGDFK-----VTPDSTTQH-NKVSISPK-----210
Db 1316 SGSDGDGRBFFDVGCGAPEDIECRSVKDPHLSLEQHQVQCDVSVGFICKNBDQFNGP 1375

QY 211 -----ITTAKPINRPSLPNSDTSKETS LTVNKETTVEKETT 250
Db 1376 FGLCYDYKIRVNCWPMQKCIITPSPPTTTPSPPTTTLPTTTPSPPT-----TTT 1430

QY 251 TKNQSTDGKEKTSAKETQSIEKTSKOLAPT SKVLAKPTPKAETTTKGPALTTPKEPT 310
Db 1431 TEPPTTSPPTTTPPTT-----LPT-----TPSPISITTTTPPTTSPPT 1474

QY 311 -----PTTKERASITPKSPPTTIKSAP-TTPKEP-----APTTKSAP-TTPKEPA 356
Db 1475 TTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPASTTTLPTTTPSPPTTTPPTTTP 1534

QY 357 PTTTKEPAPTTPKEP-----APTTKERAPTTIKSAP-TTPKEPAPTTPKAPPTTP 407
Db 1535 PTTTSPPTTTPPTTSTTLPTTTPSPPTTTPPTTTPPTTTPSPPTTTPSPPTTTPPTT 1594

QY 408 KPAAPTTPKEPPTTPKBPAPTTPKBP-----APTAPKBPAPTTPKBPAPT 460
Db 1595 TBPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPT 1654

QY 461 TPKAPAPTTPKBPPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPK 517
Db 1655 TTPSPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1714

```

RESULT 8  
Q7KTF6  
ID Q7KTF6

PRELIMINARY; PRT; 1761 AA.

QY 518 EPAPTTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 572  
Db 1715 PSSTTPSPSP-PTTMTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSPPTTTPSP 1773  
QY 573 KETAPTTPKPLTP-----585  
Db 1774 STTPTTP--CVPLCNWTKWLDGSKNFHKPGDGTDELIGVCGPGWMAANTSCRA 1831  
QY 586 -----TPEEAPPT-----TPEEAPPT-----TPEEAPPT-----TPEEAPPT 597  
Db 1832 PIGQLGQTVVCDVSGLICKNEQKPGGVIPIAFCLNVEINVQCECVTQPTMTTITTE 1891  
QY 598 KPAPTTPEELAPTTPPEPTP-----TPEEAPPT-----TPEEAPPT-----TPEEAPPT 626  
Db 1892 NPETPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1951  
QY 627 ---TPKAAAPNTP-----KEPAPTTPKEP-----APTTPKEPAPT-TPKETA 664  
Db 1952 TVTPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2011  
QY 665 PTTPKGTAPTTLKEPAPTTPKBPAPKELAPTTTKEP-----TSTTSKPA 709  
Db 2012 TTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2071  
QY 710 PT-----TPKGT-----APTTPKEPAPT-----726  
Db 2072 PTPPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2131  
QY 727 ---TTPKEPAPT-----TPKGT-APT-APT-----LKEPAPT 752  
Db 2132 TTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2191  
QY 753 PKKAPKELAPTTKGP-----TSTTSKPA-----TTPKEPAPT 788  
Db 2192 TQPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2251  
QY 789 TPKEPAP-----TTPKBPAPT-----TPEEP-----PT 812  
Db 2252 VTPPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2311  
QY 813 TSEVS-----TPTTKEPTTIHKSPDESPELSABPTPKALENSPKEP-----855  
Db 2312 TTPPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2368  
QY 856 --GVPT-TKTPAATKPEMTTAKDKTTERDLR---TTP-ETTAAAPKMT---KETATTT 904  
Db 2369 PTPPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2428  
QY 905 EXTTESKI-----TATTQVTSIT-----TQDTTPEFKITT-----934  
Db 2429 PTTTPTTPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2488  
QY 935 -----LKTTTLAPKAVTTT-KKTIITTEIMNKPEETAKPKDRATNSKATTP-----978  
Db 2489 TGTQPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2548  
QY 979 -----KPOKPTKAP-----KKPTSTKKPKTMPRVRKPTTTPRKMSTW 1018  
Db 2549 TTVTPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2608  
QY 1019 PELNPTSRIAEAMLOTTT-----RNOQTPNSKLVNPKSEADG-----1057  
Db 2609 PTTTP-----ITTTTPTTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPT 2660  
QY 1058 -----GAGETPHMLLRPHVFMPEVTP 1079  
Db 2661 TTTTPTTPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 2695



QY 475 -----PTTPKEPA-----PTTKSA-----PT-----491  
Db 605 SKTTTHTAEPATKKTTHPTTQKSTTLRIETPTTRKSTAKTTREPTTKRETTERT 664  
QY 492 -----TTKEPA-----PTTKSA-----PTTKKP 511  
Db 665 QEPSTKTTTHTAEPATKKTTHPTTQKSTTLRIETPTTRKSTAKTTREPTTKRET 724  
QY 512 SPITTKPEA-----PTTPK-----EPA-----PT 530  
Db 725 TERTTKEPTTKTTTHTTBEPTTKTTTHTTSTTLKTEPTTRKSTTKTTREPT 784  
QY 531 TPKKAPATTKPEA-----PTTPKEPATTKKPA-----PTAPK-----EPA 568  
Db 785 TKRKTERTTKEPTTRKTTTHTTKEPTTNK-----TTTKTHTTPTTKKSTTLKPTTEPT 840  
QY 569 -----PTTPKETA-----PTTKKLT-----584  
Db 841 TRKSTTKTTRPTTKRETTERTTQEPSTKTTTHTTAEPATKKTTHPTTQKSTTLRI 900  
QY 585 ---PTT-PEKLAPTPEKAPTPEELAPTTPEPT-----PTTPE-----EPA- 624  
Db 901 TEBPTTRKSTAKTTRE---PTTKRETTERTKEPTTRKTTTHTTKEPTTKTTTHTT 957  
QY 625 -----PTTPKAAENTPKEPAPTTPKEPATTPKEPA-----PTTPK 661  
Db 958 KKSSTLKPTEPTTRKSTTKTTRE---PTTKRKTERTTKEPTTRKTTTHTTKEPTTNK 1015  
QY 662 ETA-----PTTPKGTA-----PTTLKEPATTPKPAKELA-----PTTKPTS 702  
Db 1016 TTTTKTHTTPTTKKSTLKEPTTRKTTTHTTKEPTTRKTTTHTTKEPTTKTTTHTT 1075  
QY 703 TTSKPEA-----PTTPKGTA-----PTTPKEPATTPKEPA-----PTTP 737  
Db 1076 ETTAEPATKKTTHPTTQKSTTLRIETPTTRKSTAKTTREPTTRKRETTERTTKEPTTR 1135  
QY 738 KGAPTTLKEPA-----PTTPK-----KPAPKELAPTTPKPTSTSDKPAPTTPKETA 786  
Db 1136 KTTTHTTKEPTTKTTTHTTKEPTTKSTTLKPTES---PTTRK---TSITTKTREPTTKRKT 1190  
QY 787 PTTPKEPA-----PTTPK-----KP-----APTTPET 808  
Db 1191 ERTKEPTTRKTTTHTTKEPTTNKTTTHTTPTTKSTTLKPTTEPTTRKSTTKTT 1250  
QY 809 PPTTSEVSTPTTKPE---TTIHKSPDE---STBELSAEPTPKALENPKPGVP 858  
Db 1251 REPTTKRVTTERTTREPTTRKTTTHTTKEPTTKTTTHTTKEPTTKKTTTHTTPTTK-SITLKPTEEP 1309  
QY 859 TTKTPAATKEMTTAKDKTTERDLR-----TTPE 888  
Db 1310 TRKSTTKTTREPTTRKRVTTERTTREPTTRKTTTHTTKEPTTKTTTHTTKEPTTKK 1369  
QY 889 TTTAAP-----KWKETA---TTEKTESKITATTTQVSTTTQDPTTFFKIT 933  
Db 1370 STTLKPTEPTTRKSTTKTTREPTTRKRVTTERTTREPTTRKTT---THKTTREPTTKKT 1427  
QY 934 TLKT-----TLAP-----KVTTK-----KTIITTEIMNKP-----EE 962  
Db 1428 TKKTTHTTPTTKKSTLKPTEEPTRKSTTKTTREPTTRKRVTTERTTREPTTRKTTTHTT 1487  
QY 963 TAKPKDRATNSKATPKQO-----KP-----TKAPKKT-----991  
Db 1488 TEBPTTKTTTHTTPTTKKSTTLKPTTEPTTRKSTTKKATREPTTRKRVTTERTTREPT 1547  
QY 992 -----STPKPKMPVRVKPKTTPTRKMTSTMPELNP 1023  
Db 1548 TTRKTTAHTTBEPTTKTTTHTTKEPTTKSTTL-----KTEPTTRKSTTKTTREPT 1603  
QY 1024 TSRIAEAMLQTT-----TRPNQTPNSKLVEVNPKSEDAGAEGET-----1063  
Db 1604 TTRETSTVKTTADQTTKRTTAEMSTNQEPSTVETITNNSQNSQNTTSTTTEBQVHH 1663

QY 1064 -----PHMLLRPH-----1071  
Db 1664 HHHHHYHKPADLGPSTILPDLPLPLPWPPLPEILPPLPPLPPLPPLPPL 1723  
QY 1072 -----VFMEPTV-PDMYDLPVNP 1089  
Db 1724 PPLPEVNLTALISLPEISLNPPLPPLPN 1752  
RESULT 9  
AAS64673 PRELIMINARY; PRT; 1761 AA.  
ID AAS64673  
AC AAS64673;  
DT 01-APR-2004 (TREMBlrel. 27, Created)  
DT 01-APR-2004 (TREMBlrel. 27, Last sequence update)  
DT 01-APR-2004 (TREMBlrel. 27, Last annotation update)  
DE CG33300-PA (Fragment).  
GN CG33300.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hoekins R.A., Galle R.P.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
RA Abail J.P., Aghayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
RA Foslter C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Host J.M., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodagef, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou G., Zhou S., Zhu H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster."  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scherer S.E., Myers S.E., Gibbs R.A., Rubin G.M.;







```

QY 725 -----APTP-----KEPAPTPKGATTLKEPAP----- 750
Db 2128 TTRTPTRTSTDSTWSSMSASTPEPSTTGTTRTTRTPTDSTWSSMSASTTP 2187
QY 751 -----TPKKPAPKELA-----PTTKGPTSTISDKAPT-----TPK 783
Db 2188 GTTRTPTRPTSTESTDWTMSASTPEPSTTGTTRTTRTPTDSTWSSMSASTPE 2247
QY 784 E-----TAPTTPKEPAP-----TPKKKPAP----- 803
Db 2248 PSTTGTTRTTRPTPTDWTMSMSASTPEPSTTGTTRTTRTPTDSTWSSMS 2307
QY 804 -----TPE----- 807
Db 2308 SASTPESTTPELLOQDHLPIQCRQVCRQLRPHRLVOPELLQDHLPIQCRQ 2367
QY 808 -----TPPTT-----SE 815
Db 2368 PLLRSHRLHVOPELLQDHLPLSTDTWSSASTPEPSTTGTTRTTRTPTDSTMS 2427
QY 816 VSTPTTKPTTHKSP-----DESTELSAEP----- 843
Db 2428 ASTPESTTGTTRTTRPTSTESTDWTMSASTPEPSTTGTTRTTRTPTSTESTD 2487
QY 844 -----TPKALENSKPEP-----GVPTKTPAATKPEMTTAKKTTTERTDRTTP 887
Db 2488 TWSSASTPEPSTTGTTRTTRPTSTESTDWTMTAST-PEPSTT-PDTTTRTPTPT 2545
QY 888 ETTTAAPKMTKATTT-----TEKTTESKITATT 916
Db 2546 TDTTMSASTPEPSTTGTTRTTRPTPTDWTMSASTPEPSTTGTTRTTRTPTST 2605
QY 917 TQVSTTTQDTPFKTTLTKTTLAP-----KVTIT 947
Db 2606 TTMSSMSASTPEPSTTGTTRTTRPTPTDSTWSSMSASTPEPSTTGTTRTPTST 2665
QY 948 KKTITITTEIMNKPEETAKPKDRATN-----SKATTPKQ----- 981
Db 2666 ESTDWTMSASTPEPSTTGTTRTTRPTSTDSTWSSMSASTPEPSTTGTTRTPT 2725
QY 982 KPT-----KAPKPTSTK-----KPKTMP-----RVRKP 1005
Db 2726 RPTPTDSTWSSMSASTTGTTRTTRPTSTESTDWTMSASTPEPSTTGTTRTPT 2785
QY 1006 KTTPTPRKMTSTM-----PELNP-TSRIAEMLOTTTRPNQTPNSKLVEVNPKEBAG 1057
Db 2786 RPTPTDSTWSSMSASTPEPSTTGTTR-----TTPTRPTDSTWSSMSAST- 2835
QY 1058 GAEGETPHMLLRHVFMPEVTP 1079
Db 2836 -----TPE-----PSTTP 2843

RESULT 11
C6SSE6
ID Q6SSE6 PRELIMINARY; PRT; 3409 AA.
AC Q6SSE6;
IT 05-JUL-2004 (T-EMBLrel. 27, Created)
LT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
LE 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
EN plus agglutinin.
CN Name:SAGI;
CS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
CC Chlamydomonadaceae; Chlamydomonas.
CX NCBI_TaxID=3055;
EN [1]
FP SEQUENCE FROM N.A.
FA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
EA Goodenough U.W.;
EL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DB EMBL; AY450930; AAS07044.1; -

```

```

DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pistol_extensin.
DR PRINTS; PRO1222; ATROPHIN.
DR PRINTS; PRO1218; PSTILEXTENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 20.3%; Score 1185.8; DB 2; Length 3409;
Best Local Similarity 15.1%; Pred. No. 1.5e-14;
Matches 345; Conservative 143; Mismatches 411; Indels 1382; Gaps 60;

QY 1 MAWKTLPI-----YLLLLSV----- 16
Db 50 LAWVTIPVSDALQFPVNVSVTSSAATDAPSAIAYIGALLDLVWVWPQNCINDTTRSYD 109
QY 17 ----- 16
Db 110 APWSPRCALPAVQGYGYDETYIDATVYSGNSLRPFYSTCRYPASSDPYAFLEPYSMPWD 169
QY 17 -----FVIQVSSQ-----ELSCGR-----CPE 35
Db 170 DFGIQQVPLDGDMMWGFVIVVDSNLKWIDPAWTAQAGAWLGQDPRDAVWVFGTHYCSW 229
QY 36 SFERGECDCDAQCKYD----- 53
Db 230 PFV---EC---SSCENYDIADPYDPDKIANGIVPAVITALDFENASLYIYILDFGMPAG 283
QY 54 -----KCCPD----- 58
Db 284 SLDNVWYLNAYNFIGGPLPANLPTLLPSLOHALDHCRACTPDVGRGTASLQYGMQYPS 343
QY 59 ---YE----- 60
Db 344 GQPEYECSDGVTCSDGTEYVIGMIPDEWGDAAAASDSSLPWANLRTVRLSNQALYG 403
QY 61 -----SFCAEVN----- 68
Db 404 PIPEGLRSASSISSWRLQGNTELCGLPEFAAPINSLLYGLTKLGTHTVWHDPAHQ 463
QY 69 -----PTSPSS-----SKAPPPSGASQTI-----KST 91
Db 464 GGECLAPPFPSPSPRPPLPPLPSPFPVPPPPSPFPSPFPSPFPSPFPSPFPSP 523
QY 92 TKRSPPPNKKTKKVIIESEITEE-----H 117
Db 524 PELPFPSPFPPTVARGIQVGGICDPSMPSPRPFPSPFPFPFPFPFPFPFPFPFP 583
QY 118 SVSENGESSSSSSSSSSSTIWKIKSKNSAANRELQKKLVKDKNKNRTKKKTPKPV 177
Db 584 PPSPDSPPASS-----VPPSPPEP----- 602
QY 178 VDEAGSLDNGDFKVTTPTDSTTQHNKVSISPKITTAKEINPRPSLPNSDTSKETS 237
Db 603 -----SPKPPSPAPPSPAPPSPFP----- 620
QY 238 NKETTVEKETTNTKQSTGDKETTSKAKTOSIEKTSKAKLAPTSKVLAKTPKAETT 297
Db 621 -----PSPAPPSPAPPSP----- 633
QY 298 TKGPALTTKPEPTPTTKPEPAST-----TPKEPTTTIKSAPTTKPEPAPTTKSAPT 350
Db 634 ---PAPPSPQPPSPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFP 690
QY 351 TKPEPAPTTKPEPAPTTK-----EPAPTTKPEPAPTTKSAPTTKPEPAPTTKPEAPT- 405
Db 691 SPAPPLPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFPSPFP 750
QY 406 -----TPKEPAPTTKPEPTPTTKPEPAPTTKPEPAPTTKPEAPT-----APKK 448
Db 751 PSPEPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 810
QY 449 PAPTTPKPEPAPTTKPEPAPT-----TTKEPSPPTPK 479
Db 811 PEPSPSPAPPSPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFPFP 870

```

```
QY 480 EPAPTTTTSAPTTPKEAPTTTTSAPTTPKEPSPTTTPKEAPTTTPKEAP----- 529
Db 871 PPSPEPPSPAPPSLEPPSPAPPSAPPSPEPPSP-APPSAPPSQPPSPPEPPSP 929
QY 530 -----TTPKAPAPTTPKEAPTTTPKEAPTTTTPKAPAPTAPK----- 565
Db 930 PPSPEPPSPAPPSPEPPSPAPPSAPPSAPPSPTPPSPVPPSPAPPSPEPPSPAPP 989
QY 566 -----EPAPTTPKETATTPKLLTPTPKEAPTTTP-APAPTTPPELAPT----- 610
Db 990 SPDPSPAPSPAPPSNPPSPVPTTPESE-GPPSPPEPPSPAPSPPTTPTTSPPPPPPE 1048
QY 611 -----TPEPTPTTPEEAPTTTPKAAAPN-----TPKEAPTTTPKEAPTTTPK----- 653
Db 1049 PPSPEPPSPAPPLPPSEPPSPAPPSAPPSAPPSQPPSPAPPSAPPSAPPSPEPP 1108
QY 654 EPAPTTPKETAPTTPKGTAPTTL-----TKERTSTTSKAPATTP 713
Db 1109 SPAPPSPPAPPSPEPPSPAPLLPPSPDPSPAPPSMPPLPTSPSPPEVPPTPPPS 1168
QY 678 EPAPTTPKPKAPKELATT-----TKERTSTTSKAPATTP 713
Db 1169 PPAPPSAPPSQPPLEPPSPAPPSAPPSAPPSAPPSPEPPSPPEPPSPAPPSPP 1228
QY 714 KGTAPTTPKEAPTTPKEAPTTKGTAPTTLKEAPTTTPKAPKELAPT-TTKGTST 772
Db 1229 APPSP-APPSAPPSAPPSAPPSQPPS-----PAPPSPEPPSPAPPSAPPSAPPSPP 1283
QY 773 TSKAPAPTTPKEAPTTPKEAPTTTPKAPTTTTPPTT-TSEVSTPTTTPKEPTTHKS 831
Db 1284 APPSPAPPS-APPSPEPPSPAPPPSPVPPSPAPPSPTPPAPAPAAALPLPPSPA 1340
QY 832 PDSTPELSAPPTKALE-----NSPKPEGV-----PTTKTP----- 863
Db 1341 PPLVPSPAPSPSPLRPPQPTPAMPSPAPPSAPPSAPSPAGVPVPPPTTTPPLAP 1400
QY 864 -----AATKPE-----MTTAA-- 874
Db 1401 LPPDCTLLAQAALLSIPDAANSFVVSAGLPTSVAPSTPELLASCTVSCQLTATAIS 1460
QY 875 -----KDKTTERDL-----RTTP----- 887
Db 1521 AEWAVDAVQGTQQLQLSIGGVYTRTVVVDRTPPSVGNVTLSANRIKQEPSAVGEASL 1580
QY 888 -----ETTTAABKMT-----KETAT 902
Db 1581 NALGSKQAMLLTISFSEPPVPAPDPAASLIVTGALVAEWAADKMTFYVLAMTLPAELVA 1640
QY 903 TTEKTTESKIT----- 913
Db 1641 TAAGSSSSGTSRSGNGTGAATAAAAAAAPPAGTTGRRRALQQOAAAPPPASGSSSL 1700
QY 914 -----ATTQV 919
Db 1701 SGAATANQOQRHVHFLPATAYADAARNPGRNDLSLSVELTDNAVASPAVGEALATTAR 1760
QY 920 TSTTTQDTPPFKITTLTKTLTAPKVT----- 945
Db 1761 TAATYP-----AAVATTLVAASSSFAQAIKAGSLLOGSYHTQMLTMSLYLASRGV 1813
QY 946 -----TTKKTITTTT-----IMNK 959
Db 1814 GREYGEYAVEFKYAVLKGKNGLGAFAEAAMPTNEKEVTAABQARQVGGDLWPIGNLILGGS 1873
QY 960 PBETAKPKORATNSKATTPKPKP----- 983
Db 1874 SNTTASGSSGSSSSNSP-PRRPPPPPAAGTGLLFSNADASPPPLAVATPALPAPLPS 1932
```

## RESULT 12

AAS07044

```
ID AAS07044 PRELIMINARY; PRT; 3409 AA.
AC AAS07044;
DT 02-MAR-2004 (T-EMBLrel. 27, Created)
DT 02-MAR-2004 (T-EMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (T-EMBLrel. 27, Last annotation update)
DE Plus agglutinin.
GN SAG1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waffenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RT Goodenough U.W.;
RA "Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AAS07044.1;
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633BDD6631F CRC64;
```

Query Match 20.3%; Score 1185.8; DB 2; Length 3409;  
Best Local Similarity 15.1%; Pred. No. 1.5e-14;  
Matches 345; Conservative 143; Mismatches 411; Indels 1382; Gaps 60;

```
QY 1 MAWKTLPI-----YLLLLLSV----- 16
Db 50 LAWVTIPVSDALQFPVNVSVTSSAATDAPSAIYIGALLDLVWNPQPNCTNDRSYD 109
QY 17 ----- 16
Db 110 APWPSRCALPAVQGYGYDEYIDATYVSGNSLRPFYSTCRYPASSDPYAFLEPYMPWVD 169
QY 17 -----FVIQVSSQ-----ELCKGR-----CPE 35
Db 170 DFGIYQPVLDGDFDMWGWFTVNTVDSNLKWIDPAWTAQGAWLGGQDPRDVAWNEGTHYCSW 229
QY 36 SPERGECDDAQCKKYD----- 53
Db 230 PFV---EC---SSCENYDIADVPDPDKIANGIVPAVITALDPFNASLIYIYLDFGMFAG 283
QY 54 -----KCCPD----- 58
Db 284 SLDNVWYLNAYNFTGGPLPANLPTLLPSLOHALDHCRAETDVRGTASLQYGWQYQYPT 343
QY 59 ---YE----- 60
Db 344 GQPYECSDGVTGSDGTEYVIGMIPDEWGDVAASDSSLPWANLFTVLSNQALY 403
QY 61 -----SPCAEVHN----- 68
```



SEQUENCE FROM N.A.  
 RC STRAIN=NCIB 10682;  
 RX MEDLINE=93209931; PubMed=8459832;  
 RA Fujino T., Beguin P., Aubert J.-P.;  
 RT "Organization of a Clostridium thermocellum gene cluster encoding the  
 RT cellulosomal scaffolding protein Cipa and a protein possibly involved  
 RT in attachment of the cellulosome to the cell surface.";  
 RL J. Bacteriol. 175:1891-1899(1993).  
 CC -/- SUBUNIT: Assembled into mono-layered crystalline arrays.  
 CC -/- SUBCELLULAR LOCATION: Cell wall.  
 CC -/- SIMILARITY: Contains 4 S-layer homology (SLH) domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; X67506; CAA47841.1; -;  
 DR PIR; T18262; T18262.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR001119; SLH.  
 DR Pfam; PF00335; SLH; 3.  
 DR PROSITE; PS01072; SLH\_DOMAIN; 2.  
 KW Cell wall; Repeat; S-Layer; Signal.  
 FT SIGNAL 1 28 Potential.  
 FT CHAIN 29 1664 Cell surface glycoprotein 1.  
 FT DOMAIN 36 763 4 X 156 AA approximate repeats.  
 FT REPEAT 36 191 1.  
 FT REPEAT 207 363 2.  
 FT REPEAT 409 565 3.  
 FT REPEAT 607 763 4.  
 FT DOMAIN 771 1377 approximate tandem repeats of T-P-S-D-E-P.  
 FT  
 FT DOMAIN 1378 1449 Gly/Pro/Ser/Thr-rich.  
 FT DOMAIN 1453 1494 SLH 1 (incomplete).  
 FT DOMAIN 1495 1565 SLH 2.  
 FT DOMAIN 1566 1625 SLH 3.  
 FT DOMAIN 1626 1646 SLH 4 (incomplete).  
 FT SEQUENCE 1664 AA; 178194 MW; 5F396695BA9FE74B CRC64;  
 SQ  
 Query Match 19.6%; Score 1143.3; DB 1; Length 1664;  
 Best Local Similarity 21.2%; Pred. No. 4e-14;  
 Matches 367; Conservative 131; Mismatches 425; Indels 807; Gaps 66;  
 QY 4 KTLPIVILLLLSVFV-----IQVSSQE 26  
 DB 6 KVLISILLTLLIISTTSVNMFAEATPSIEWVLDTKEVHVGDVITATIKVNNIRKLAGYQ 65  
 QY 27 LSKGRCFBSFERGRCDCDAQCKYDKCPDYSECAEVHNTPSPSSK---KAPPS 82  
 DB 66 LNIK-----FDPEVLQVDPATGEEFTDKSMP-- 92  
 QY 83 GASQIKSTKRSPPNKKTKKVISEBITEHSHVSENOESSSSSSSSSSSTTWKIK 142  
 DB 93 -VNRVLLTNSKYGPT-----VAGNDIKSGIINFATGNNLTAYK 131  
 QY 143 SS----- 144  
 DB 132 SSGIDEHTGIIGIFKVLKQNTSIRFEDTLSPGALSGTSLFDWDAETITGYEVIQPD 191  
 QY 145 -----KNAANRELQK-KLVKD-----NKKN-----RTKKKT----- 172  
 DB 192 LIVVEAPLKASVALELDKTKVKGDIITAIKIKENMKNFAGYQINIKYDPTMLEAIEL 251  
 QY 173 -----PKPEVVDAGSLDN----- 187  
 DB 252 ETGSAIAKRTWPTGTVQLSDNYGKTTAVANDVGAGIINFABEAYSNLTKYRTGVAET 311  
 QY 188 -----GPKVTTPTD-----STQHNKV----- 205

Db QY 312 GIIGKIGFRVLKAGSTAIRFEDTTAMPGAIEGTYMFDWYGENIKGYSVVQGEIVAESEE 371  
 QY 206 -----STSPKITTAKPINRPSLP-----PNS-----DTSK----- 231  
 Db 372 PGESEPTPEPVTETPVDPTPTVTEEPVPSLPSYVIMELDKTKVKVGDIIITAIKIKENM 431  
 QY 232 -----ETSLTVNKETTVETKETTTTKTKQSTDKGKETS----- 265  
 Db 432 KNFAGYQINIKYDPTMLEALETGSIAKRTWPVTGGTV-----LOSDNYGKTTAVAND 486  
 QY 266 -----AKETQSIEKTSKADLAPTSKVLAKEPTPKAETTT----- 298  
 Db 487 VGAGIINFABEAYSNLTKYRTGVAETGIIGKIGFRVLKAGSTAI-----RFEDTTAMPG 541  
 QY 299 -----KGPALTTTP-----KEPTPT-TPKEPASTTPKEPTPT----- 328  
 Db 542 AIBGTVMFDWYGENIKGYSVVQGEIVAEGEETPEPVTETPVDPTPTVTEEPVPSLPS 601  
 QY 329 ----- 328  
 Db 602 DSVVIMELDKTKVKEGDVIIATIRVNNIKNLGAYQIGIKYDKPVKLEAFNIETGPDIDEGT 661  
 QY 329 ----- 329  
 Db 662 WPAVGGTILKNRDYLPVGAINNVSKILNFAAYVYVFDYDREKSESDTGIIIGNIGFRV 721  
 QY 330 IKSAPITPK-----EPATTTKSAPTTPKEPAPTT 359  
 Db 722 LKAEDTTIRFEELSPGSDGTGYMLDWNLRISGVYVQIPAKAAS-----DEPIPTD 776  
 QY 360 TKEPAPTTTPKEPAP-----TTTKEPAPTTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAP 413  
 Db 777 TPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 836  
 QY 414 -----TPK-EPTPT-TPKEPAPT-TKEPAPTTTPKEPAPTAPKKA-----PTTTPKEPAPT- 460  
 Db 837 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 896  
 QY 461 TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKSAPTTTTPKEPAPTTTPKSAPTTTTPKEPAP 520  
 Db 897 TPEEPITPTPSDEPTPSDEPTPS-----DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 953  
 QY 521 PTTTPKEPAPTTTPKKA-----EPTTPKEPAPT-TPEKEPAPTTTPKKAAPTAPK 565  
 Db 954 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1013  
 QY 566 EPAPITTPKSAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 621  
 Db 1014 EPTPSD-----EPTPSDEPTPSDEPTPSD-EPTPSDEPTPSDEPTPSDEPTPSDEPT 1052  
 QY 622 EPAPITTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 678  
 Db 1053 EPIPTPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1112  
 QY 679 PAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 737  
 Db 1113 PTPS-----DEPTPSD-EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1165  
 QY 738 KGTAPTTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 793  
 Db 1166 --DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1217  
 QY 794 APT-TPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTP 952  
 Db 1218 TSEPTPSDEPTPSDEPTPSD-EPTPSDEPTPSD-EPTPSDEPTPSD-EPTPSDEPTPSD-EPT 1266  
 QY 853 KEPGVPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPTTPKKAAPT 912  
 Db 1267 EEP-IPDTTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1325  
 QY 913 TATTQVSTTTQDTTPFKITLTKTTTLAPKVTTTKKTIITTEIMNKPBETAKPKDRATN 972  
 Db 1326 PTDTPSDEPTPSDEPTPSDEPT-----PSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1360

```

QY 973 SKATTPKQ-KTKAPKPTSTKKXTMPRVKPKTKTTPRKMVS----- 1016
DQ 1361 SDEPTSEDEPTSEPTPEPT-----PTTPTPTPTSTTPSGSGSGGGG 1407
QY 1017 -----TMPELNPTSRLAEAMLQTTTRPNOTPNKSLVENPKSBDAGAGETPHMLL 1068
DQ 1408 GGGGTVPTSPPTPTIS-----KPTSTPAPTEIE-BPTSDVPGAIGGEHAYL 1455
QY 1069 RPH-----VFMPE----- 1076
DQ 1456 RGYPDGSPERNITRAEAAVIFAKLLGADESYGASPSYDLADTHWAAWAIKPAISQ 1515
QY 1077 -----VTPD-----MDYLPVPNQG-----IINP 1096
DQ 1516 GLFKGYDGTFRPDQNIIRAFAVVLHFLTKVKGQEIIMSKLATIDISNP 1565

RESULT 14
Q3WWQ4 PRELIMINARY; PRT; 1349 AA.
AC Q8WWQ4;
DI 01-MAR-2002 (TrEMBLrel. 20, Created)
DI 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DI 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mucin 5 (Fragment).
GN Name=MUC5AC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RV [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21426417; PubMed=11535137;
RA Escande F., Aubert J.P., Porchet N., Buisine M.P.;
RT "Human mucin gene MUC5AC: organization of its 5'-region and central
RT repetitive region.";
RL Biochem. J. 358:763-772(2001).
DR EMBL; AJ298318; CAC83675.1; -.
FT NON TER 1
FT NON TER 1349 1349
SQ SEQUENCE 1349 AA; 135599 MW; 4DC3C1544F1E5EBA CRC64;

Query Match
Best Local Similarity 19.1%; Score 1117.1; DB 2; Length 1349;
Matches 355; Conservative 73; Mismatches 371; Indels 709; Gaps 44;

QY 74 SSKKAPPPGASCTIKSTTKRSPKPKPKKKTKKVIIESEITEHSEHSVSENOESSSSSSSS 133
DQ 3 STTSAPTTSTTSTSTST-----STPQTSTTSASTIS 35
QY 134 SSGTIWKISSKNSAANRELQKKLVKDNKNRTKKKP--TPKPPVVVDEAGSLDNGDFK 191
DQ 36 ITS-----GPGTTPSP----- 46
QY 192 VITPDTSTTQHNKVSIPKITTAKPINRPSLPSNDTSKETSILVNKETTIVTKETTTT 251
DQ 47 --VPTTSTT-----SAPTTSTSAATSTTSAPTSTTSAPTTS 83
QY 252 NKQTSIDGKEK-----TTSAKETQSTEKTSKADLAP-----TSKVLA 288
DQ 84 TTSASTASKTSGLGTTPSPPIPTTSTSPPTTSTTSASTASKTSGPGTTPSPVPTSTIPA 143
QY 289 KPTPKAET-----TTKGPAITPKPEPTTT----- 313
DQ 144 ---PRSTTSASTSTTPGGG-IIP-SPVPTTSTASVSKTSHSVSISKITHSQPVTRDC 198
QY 314 -----PKE----- 316
DQ 199 HLRCTWTKWFDVDFPPSPGPHGGDKETNNIIRSGEKICRRPQETIRLQCRAKSHPEVSIE 258
QY 317 -----PASTTP----- 322

```

```

DQ 259 HLGQVVQCSREBGLVCRNQDQGPFKMCLNVEVRVLCCETPKGCPTSTPTVAPSTPSGR 318
QY 323 -KEPT-----PTTIKSAPTTKEPAPTT-TKSAPTTKEPAPTTTKEPAP 365
DQ 319 ATSPTQSTSSWOKSRITTLVTSTSTPQTSTTSAPTSTTTSAPTSTTSTPT 378
QY 366 TTPKEPAPTTTKEPAPTTIK-----SAPTTKEPAPTTTKEPAPTTTKEPAPTTTPK 416
DQ 379 QTSISSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTST 438
QY 417 EPTPTTKEPAP-----TTKEPAPTT----- 437
DQ 439 TSTPQTSKTSAAATSTSTSSGTTTPSPVTTTSTASVSKTSHSVSVSKTTHSQPVTRDCHP 498
QY 438 -----PKE----- 440
DQ 499 RCTWTKWFDVDFPPSPGPHGGDKETNNIIRSGEKICRRPQETIRLQCRAKSHPEVSIEHL 558
QY 441 -----PAPTAPKKP 449
DQ 559 GQVVQCSREBGLVCRNQDQGPFKMCLNVEVRVLCCETPKGCPTSTSTVAFSLVGEPP 618
QY 450 APT-----TPKEPAPTTTKEPAP-----PTTKEPSTTTPKEPAP 483
DQ 619 AQTQSTSSWOKSRITTLVTSS:ITSTQSTSTTSAPTSTTTPASIPSTTS--APTSTTSAP 676
QY 484 TT-TKSAPTTTKEPAP-TTTSKAPTTKEPSTTTSAPTSTTKEPAPTTTKEPAPTTTKEPAPTTTPK 541
DQ 677 TSTTSAPTSTTSTPQTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTST 736
QY 542 EPAPTTKEPAPTTTTPKAPPTAPKAPPTTKEPAPTTTTPKAPTTTTPKAPTTTTPKAPPT 601
DQ 737 TSAPTSTSS-APTNTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSTPTPT 795
QY 602 TTPPELAPTTPEEPTTPEEAPTTTTPKAAAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTPK 661
DQ 796 QSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 855
QY 662 ETAPITPKGTAPTTTKEPAPTTK-----KPAPKELAPTTTKEPAPTTTSTSTSTSTSTSTSTST 705
DQ 856 TSFHTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 912
QY 706 -----DKPAP----- 710
DQ 913 HVSVKITHSQPVTRDCHPRCTWKFVDVDFSPGPHGGDKETNNIIRSGEKICRRPPEE 972
QY 711 -----TPPKG 715
DQ 973 ITRLQCRAESHPESVIEHLGQVVQCSREBGLVCRNQDQGPFKMCLNVEVRVLCCETPKG 1032
QY 716 -----TAPTPK-----EPAPTTKEPA 733
DQ 1033 CPVTSTPTVAPSTPSGRATSPQTSTSSWOKSRITTLVTSTSTPTSTPTSTSTSTSTSTPTIPA 1092
QY 734 PTPPKGTAPTTTKEPAPTTTTPKAPKELAPTTTTPKGTSTSTSDKPAPTTTPKAPTTTTPKAP 793
DQ 1093 STPSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTSAPTST 1147
QY 794 APPTPKKAPPTTPEPTPTTSEVSTPT--TTKEPPTTIHKSPODESTPELSAEPPTPKALENS 851
DQ 1148 APPTSTISASTSTISAPTSTSTISAPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1202
QY 852 PKEPGVPTTKTAAKPEMTTAKOKTTERDLRTPETTTAAPKMTKEATITTEKTESK 911
DQ 1203 P-----VPTTSTTSAPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1253
QY 912 ITATTTQVSTTTQ--DTPPTTKITTLTKTTLAPKVTTTKKTTTITTEIMNKPETEAPKXDR 969
DQ 1254 TTRTTSAPTSSMTSGGTTTSPVPTTSTTS-APTST----- 1289
QY 970 ATNSKATTPKPKQKPTKAPKKP-TSTKKPKTMPRVKPKTKTTPTRKMTSTMPELNPTSRIA 1028
DQ 1290 -TSGPGTTPSPVPTTSTTSAPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1338

```

QY 1029 EAMLOTT 1036  
 Db 1339 SASTATT 1346

RESULT 15  
 Q81R51  
 ID Q81R51 PRELIMINARY; PRT; 3432 AA.  
 AC Q81R51;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE CG32602-PB.  
 GN ORFNames=CG32602;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Fertaz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasse K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris D.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin N., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhang M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Murny D.M., Nelson C.R.,  
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirkas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 melanogaster euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).

[3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirkas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the Drosophila melanogaster euchromatic  
 a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 systematic review.";  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX FLYBASE;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX FLYBASE;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 RG EMBL; AE003495; AA09587.1; -  
 DR FLYBase; Fg00052602; CG32602.  
 SQ SEQUENCE 3432 AA; 349492 MW; FE907AC75578657A CRC64;

Query Match 19.1%; Score 1112.3; DB 2; Length 3432;  
 Best Local Similarity 26.0%; Pred. No. 3.9e-13;  
 Matches 466; Conservative 68; Mismatches 410; Indels 847; Gaps 92;  
 56 CPDYSEFCAEVHNTSP-----SSSSSSSTIWK-----IKSSKNSAAN 150  
 58 CPSPHTTPACEHRITSPAPTKPHCPHATPHNCGHDTSSSRATTPKPTLETTIGTTFK 117  
 73 -----PSSKAPPSPGAS---QIKSTKSPKPKPKKTKKVKVIESEIEHSVSE 121  
 118 TTQITTRGTTQK--PTEGSAQPTTLKTEGTAKPTLKPT-----EGTAKPTTLK 169  
 122 NOESSSS-----SSSSSSSTIWK-----IKSSKNSAAN 150  
 170 PTEGSAQPTTLKTEGTAKPTTLKTEGTSAPTKPTLKTEGTTAKPTLKTEGTTA- 228  
 151 RELQKLLKVDNKNKTKKPKPKPVVDGAGSLDNGDFKVTTPDTS----- 198  
 229 ----KPTTLKPTB--GTTAKPTLKP-----TEGSAKPTTLKTEGT 265  
 199 ----TQHNKVTSPKITTAKP-----INPRSLPPNSDTSKETSLSL----- 235  
 266 TAKPTTLKTEGTSAPTKPTLKTEGTTAKPTTLKTEGTTAKPTTLKTEGTSAPTKTLK 325  
 236 ----TVNKEETTVEKETT-----TTNKQTS-----TDGK 260  
 326 PTEGTTAKPTTLKTEGTSAPTKPTLKTEGTTAKPTTLKTEGTSAPTKPTLKTEGTS 385  
 261 EKT-----TSAKET--QSIKTSK--DLAPTSKVLAKPT---PKAETTKGPALTTPK 307  
 386 PTLKTEGTSAPTKPTLKTEGTSAPTKPTLKTEGTSAPTKPTLKTEGTTAK--PTTLKPT 444  
 308 EPT---PTTPKEPASTTPK---EPT-----PTTIK-----SAPTPKEPAPITTKSA 348  
 445 EGTSAPTKTLKTEGTTAKPTTLKTEGTTAKPTTLKTEGTSAPTKPTLKTEGTTAK-- 502  
 349 PTPPK-----EPAPITTKPA-----PTTPKEPAPITTKPEAPITTT-- 383  
 503 PTLKTEGTSAPTKPTLKTEGTTAKPTTLKTEGTSAPTKPTLKTEGTTAK---PTTLK 559



```
QY 384 ----TKSAPTKEPAPTPK---KPA-----PTTPKEPAPTPK-----BPTPTPK 425
Db 560 PTEGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTSAP 619
QY 426 PA-----PTTKEPA-----PTTPKEPAPTPKPAKPTTPK 455
Db 620 PTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLK---PTEGTSAPKPTTLK 676
QY 456 EPAPTPPK-----EPAPTTKEPS-----PTTPKEPAPTPPK 487
Db 677 PTEGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKLTGTSAPKPTTLKPTGTTAK 736
QY 488 -----SAPTTTKEPA-----PTT-----TKSAPTTPKESPPTTKEPA----- 520
Db 737 PTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLK---PTGTTAKPTTLKPT 795
QY 521 -----PTTPKEPAPTPPK-----KPA-----PTTPKEPAPTPPK-- 549
Db 796 EGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTTAKPT 855
QY 550 ---EPAPTTTKKPA---PTAPKEPAPTPKETAPTTPKLT-----PTT----- 587
Db 856 TLKPTDGTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTDGT 915
QY 588 ---PEKLAPTTPKEPAPTT-----PEELAPT--TPEEPTPTT 619
Db 916 TAKPTTLKPTGTSAPKPTTLKPTDGTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLK 975
QY 620 PEE---PAPTPPK-----AAAPNTPEPAPTPPKBAPTPPK-----EPA----- 656
Db 976 PTEGTTAKPTTLKPTGTSAPKPTTLKPTGTTAK---PTTLKPTGTSAPKPTTLKPTGTT 1032
QY 657 ---PTTPKETAPTTPKGT-----PTTLKEPAPTPK---KPA-----PKELA 693
Db 1033 SAKPTTLKPTDGTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLK 1092
QY 694 PT-----TTKEPTSTSDKPA-----709
Db 1093 PTEGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPK 1152
QY 710 PTPPKGTAPTTPKEPAPTPPKBAPTPPKGTA-----PTTLKEPAPTPPK---K 755
Db 1153 PTTLKPTGTTAK---PTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLK 1209
QY 756 PA-----PKELAPT-----TTKGTSTSDKPA-----778
Db 1210 PTEGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTDGTAK 1269
QY 779 -----PTTPKETAPTTPKEPAPTPK-----KPA-----PTT-- 805
Db 1270 PTTLKPTGTSAPKPTTLKPTGTTAK---PTTLKPTGTTAKPTTLKPTDGTAKPTTLK 1326
QY 806 -----PETPPPTTSVSTPT-----TTKEPTTIHKSPDESTP 837
Db 1327 PTEGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTDGTAKPTTL---KPTGTT- 1383
QY 838 ELSAPPT-----PKALENSPK-----PGVPTTKTPAAKPBMTTAKD 876
Db 1384 --SAKPTTLKPTGTTAKPTTLK---PTEGTSAPKPTTLKPTGTTAKPTTLKPTGTTAKP 1439
QY 877 KTTTERDLRTT-----PETTTAAPKMTKATATTEKTESKITAT 915
Db 1440 TTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTKG 1499
QY 916 T-----916
Db 1500 TSAKPTTLKPTGTTAKPTTLKPTDGTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTL 1559
QY 917 -----TQVTS-----TTTQDPTT---PFKITTLKTTTLAP----- 942
Db 1560 KPTEGTSAPKPTTLKPTGTSAPKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTS 1619
```

```
QY 943 KYTTTKKTLTTTBMNKPEETAKPKDRATNSKATTPKPOKPTKAPKPT-----STKK 995
Db 1620 KPPTTLKPTGTTA---KP---TTLKPTD-GTTAKPTTLKPTGTSAPKPTTLKPTGTTAK 1672
QY 996 PKTMPRVVRKP-----KITTPTRKMTSTMP-ELNPTSRISAZMLQTTTRPNQTPNSKLV 1047
Db 1673 PTTL---KPTEGTSAPKPTTLKPTBRTSAQPTTLKPTERTSAQ--PTTLKPTGTTAKPT 1726
QY 1048 EYNPKSEDAGGABGET--PHMLLRPHVFMPEVTPDMDYLPVRPNQGIINP 1096
Db 1727 TLKP-----TEGTTAKPTTL-----KPTEGTTAKP 1751
```

Search completed: October 13, 2004, 12:16:27  
Job time : 151.835 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 18.2347 Seconds  
(without alignments)  
3815.116 Million cell updates/sec

T.tle: SEQ1-D

Perfect score: 5576

Sequence: 1 MAWKTLPIYLLLLLSVFVIQ.....DMDYLPVNOGIINPMLS 1049

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA.\*
- 1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep.\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep.\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/iaa/PCFUS-COMB.pep.\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5576	100.0	1049	4	US-07-757-022B-58
2	5576	100.0	1313	4	US-07-757-022B-142
3	5561.9	99.7	1354	4	US-07-757-022B-48
4	5561.9	99.7	1363	4	US-07-757-022B-52
5	5546.9	99.5	1140	4	US-07-757-022B-104
6	5546.9	99.5	1404	4	US-07-757-022B-2
7	5546.9	99.5	1404	4	US-07-757-022B-62
8	5546.9	99.5	1404	4	US-07-757-022B-1
9	5523.9	99.1	1404	4	US-10-164-595-78
10	5447.9	97.7	1314	4	US-07-757-022B-50
11	5393.3	96.7	1320	4	US-07-757-022B-46
12	5393.3	96.7	1320	4	US-07-757-022B-60
13	5384.3	96.6	1320	4	US-10-164-595-58
14	5379.2	96.5	1361	4	US-07-757-022B-40
15	5358.7	96.1	1038	4	US-07-757-022B-74
16	5358.7	96.1	1270	4	US-07-757-022B-44
17	5344.6	95.9	1311	4	US-07-757-022B-42
18	5235.9	93.9	1022	4	US-07-757-022B-84
19	5011	89.9	941	4	US-07-757-022B-14
20	1346.8	24.2	5179	4	US-07-538-092-1258
21	1088.5	19.5	8991	4	US-08-714-741-32
22	827.1	14.8	3256	4	US-09-919-172-98
23	827.1	14.8	3256	4	US-09-976-594-22
24	827.1	14.8	3256	4	US-09-919-039-21
25	816.3	14.6	3118	3	US-09-579-181-1
26	809.6	14.5	2972	3	US-09-579-181-2
27	762.9	13.7	1837	3	US-08-928-361B-5

28 762.9 13.7 1837 4 US-09-588-995A-5  
29 756.1 13.6 2142 4 US-09-538-092-1142  
30 735 13.2 1721 3 US-08-700-651-5  
31 733 13.1 1721 3 US-08-928-361B-6  
32 733 13.1 1721 4 US-09-588-995A-6  
33 716 12.8 2736 4 US-09-252-991A-30227  
34 705.3 12.6 807 4 US-09-294-663-4  
35 697.5 12.5 4019 4 US-08-854-133-425  
36 692.7 12.4 3969 3 US-08-061-376-5  
37 692.6 12.4 2468 4 US-09-976-594-726  
38 692.6 12.4 2468 4 US-09-538-092-1135  
39 687.4 12.3 3969 4 US-09-538-092-1262  
40 686.1 12.3 805 3 US-09-103-429A-4  
41 679.9 12.2 788 4 US-09-294-663-3  
42 675 12.1 827 4 US-09-248-796A-17307  
43 659.8 11.8 786 3 US-09-103-429A-3  
44 649.9 11.7 220 4 US-07-757-022B-96  
45 640.2 11.5 207 4 US-07-757-022B-116

## ALIGNMENTS

RESULT 1  
US-07-757-022B-58  
; Sequence 58, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1049 amino acids  
; TYPE: AMINO ACID

**This Page Blank (uspto)**

```
Qy 121 PTPKPPVDEAGSLDNGDFKVTTPDTSTQHNKSVSTPKITTAKEINRPSLPNSDTS 180
Db 121 PTPKPPVDEAGSLDNGDFKVTTPDTSTQHNKSVSTPKITTAKEINRPSLPNSDTS 180
Qy 181 KETSLTVNKETTVETKETTITNKQTSIDGKEKTTSAKETOSIEKTSKOLAPTSKVLAKP 240
Db 181 KETSLTVNKETTVETKETTITNKQTSIDGKEKTTSAKETOSIEKTSKOLAPTSKVLAKP 240
Qy 241 TPXAETTTKGPAUTTPKBPPTTPKBPASTTPKEPTTTIKSAPPTTPKEPAPTTKSAPT 300
Db 241 TPXAETTTKGPAUTTPKBPPTTPKBPASTTPKEPTTTIKSAPPTTPKEPAPTTKSAPT 300
Qy 301 TPKEPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 360
Db 301 TPKEPAPTTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 360
Qy 361 APPTTKEPTTTPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTT 420
Db 361 APPTTKEPTTTPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTT 420
Qy 421 KPSPTTPKBPAPTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 480
Db 421 KPSPTTPKBPAPTTTKSAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPT 480
Qy 481 TPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 540
Db 481 TPKKPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 540
Qy 541 LAPTTPEKAPTTPEELAPTTPEEPTTTPPEAPTTTPKAAAPNTKPEAPTTKBPAPT 600
Db 541 LAPTTPEKAPTTPEELAPTTPEEPTTTPPEAPTTTPKAAAPNTKPEAPTTKBPAPT 600
Qy 601 TPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 660
Db 601 TPKEPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 660
Qy 661 TTPKGTAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 720
Db 661 TTPKGTAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 720
Qy 721 STTSKAPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 780
Db 721 STTSKAPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAPTTKBPAP 780
Qy 781 SPDESTPELSAETPKALNSKBPVPTTKTAPKAPKPTSTKPKTMRVRKPKTTPT 840
Db 781 SPDESTPELSAETPKALNSKBPVPTTKTAPKAPKPTSTKPKTMRVRKPKTTPT 840
Qy 841 TAAPKMTKATTTKTESKITATTTQVSTTODTTPKTIITLKTTLAPKVTTKT 900
Db 841 TAAPKMTKATTTKTESKITATTTQVSTTODTTPKTIITLKTTLAPKVTTKT 900
Qy 901 ITTTEIMNKPEETAAPKALNSKATTPKPKPTKAPKPTSTKPKTMRVRKPKTTPT 960
Db 901 ITTTEIMNKPEETAAPKALNSKATTPKPKPTKAPKPTSTKPKTMRVRKPKTTPT 960
Qy 961 PRKMTSTMPNLPTSIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRP 1020
Db 961 PRKMTSTMPNLPTSIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAGETPHMLLRP 1020
Qy 1021 HVFMPEVTPDMXYLPRVFNQGIINPMLS 1049
Db 1021 HVFMPEVTPDMXYLPRVFNQGIINPMLS 1049
```

## RESULT 3

US-07-757-022B-48

Sequence 48, Application US/07757022B

Patent No. 6433142

GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

```
APPLICANT: Turner, Katherine
APPLICANT: Hewick, Rodney M.
TITLE OF INVENTION: Megakaryocyte Stimulating Factors
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/757,022B
FILING DATE: 19910910
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Cseri, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELEPHONE: (617) 876-1170
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-48
```

```
Query Match 99.7%; Score 5561.9; DB 4; Length 1354;
Best Local Similarity 96.2%; Pred. No. 1.3e-169;
Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSACAGCGEGYSDATCNCNDYNQHYMECCPDF 60
Qy 26 -----ELSCGKCFSEFERGECDDAQCCKYDKCCPDYESFCABEHSVSENOSSSS 79
Db 61 KRVTAELSCKGCFSEFERGECDDAQCCKYDKCCPDYESFCABEHSVSENOSSSS 120
Qy 80 SSSSSSTTWIKSSKNSAANRELQKLVKDNKNKRTKKPTKPPVVDAGSLDNGD 139
Db 121 SSSSSSTTWIKSSKNSAANRELQKLVKDNKNKRTKKPTKPPVVDAGSLDNGD 180
Qy 140 FKVTTPDTSTQHNKSVSTPKITTAKEINRPSLPNSDTSKETSITVTKETTVETKETT 199
Db 181 FKVTTPDTSTQHNKSVSTPKITTAKEINRPSLPNSDTSKETSITVTKETTVETKETT 240
Qy 200 TTNKQTSIDGKEKTTSAKETOSIEKTSKOLAPTSKVLAKPTKAEITTKGPALTTPKEP 259
Db 241 TTNKQTSIDGKEKTTSAKETOSIEKTSKOLAPTSKVLAKPTKAEITTKGPALTTPKEP 300
Qy 260 TPTTPKEPASTTPKEPPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTPK 319
```

301	DB	TPRTPKBPASGTTPKBPTPTTIIKSGARTPKBPARTTTKASPTTPKBPARTTTKBPARTTPK	360
320	QY	EPARTTTKBPARTTTKSGARTTPKBPARTTTPKKQAPRTTPKBPARTTPKBPARTTPKBPART	379
361	DB	EPARTTTKBPARTTTKSGARTTPKBPARTTPKBPARTTPKBPARTTPKBPARTTPKBPART	420
380	QY	TKBPARTTPKBPARTAPKKQAPRTTPKBPARTTPKBPARTTTKBPSPRTTPKBPARTTTKSA	439
421	DB	TKBPARTTPKBPARTAPKKQAPRTTPKBPARTTPKBPARTTTKBPSPRTTPKBPARTTTKSA	480
440	QY	PTTTTKBPARTTTKSGARTTPKBPSPRTTTKBPARTTPKBPARTTPKBPARTTPKBPARTTPK	499
481	DB	PTTTTKBPARTTTKSGARTTPKBPSPRTTTKBPARTTPKBPARTTPKBPARTTPKBPARTTPK	540
500	QY	EPARTTTTKBPARTAPKBPARTTPKBPARTTPKBLTPTTPKBLARTTTPEKAPRTTPEELAP	559
541	DB	EPARTTTTKBPARTAPKBPARTTPKBPARTTPKBLTPTTPKBLARTTTPEKAPRTTPEELAP	600
560	QY	TTPEEPRTTTPKBPARTTPKKAAPNTPKBPARTTPKBPARTTPKBPARTTPKBPARTTPK	619
601	DB	TTPEEPRTTTPKBPARTTPKKAAPNTPKBPARTTPKBPARTTPKBPARTTPKBPARTTPK	660
620	QY	GTAPRTTLKBPARTTPKBPAPKELAPRTTTKBPSTSTSDKAPRTTPKGTAPRTTPKBPARTTP	679
661	DB	GTAPRTTLKBPARTTPKBPAPKELAPRTTTKBPSTSTSDKAPRTTPKGTAPRTTPKBPARTTP	720
680	QY	KEPARTTPKGTAPRTTLKBPARTTPPKBPAPKELAPRTTTKGTSTSTSDKAPRTTPKGTAPRT	739
721	DB	KEPARTTPKGTAPRTTLKBPARTTPPKBPAPKELAPRTTTKGTSTSTSDKAPRTTPKGTAPRT	780
740	QY	PKBPARTTPKBPARTTPPETPPPTSEVSTTTTKBPRTTIHKSPDESTPELSABPTPKALE	799
781	DB	PKBPARTTPKBPARTTPPETPPPTSEVSTTTTKBPRTTIHKSPDESTPELSABPTPKALE	840
800	QY	NSPKBPVPRTTKTPAATKPBMTTTAKDKTTERDLRTTTPETTTAAPKMTKSTATTTTEKTE	859
841	DB	NSPKBPVPRTTKTPAATKPBMTTTAKDKTTERDLRTTTPETTTAAPKMTKSTATTTTEKTE	900
860	QY	SKITATTTQVSTTTQDPTTPFKITTLKTTLAPKVTTTKKTIITTTTEIMNKPEETAKPKDR	919
901	DB	SKITATTTQVSTTTQDPTTPFKITTLKTTLAPKVTTTKKTIITTTTEIMNKPEETAKPKDR	960
920	QY	ATNSKATTPPKQKPTKAPKPTSTKKPKTTPRVKPKPTTTPKMTSTMPELNPTSRIAE	979
961	DB	ATNSKATTPPKQKPTKAPKPTSTKKPKTTPRVKPKPTTTPKMTSTMPELNPTSRIAE	1020
980	QY	AMLQTTTRPNQTPNSKLVEVNPXSDEAGGAGETPHMLLRPHVMPBVTTPMDYLLPRVFN	1039
1021	DB	AMLQTTTRPNQTPNSKLVEVNPXSDEAGGAGETPHMLLRPHVMPBVTTPMDYLLPRVFN	1080
1040	QY	QGIHINPMLS 1049	
1081	DB	QGIHINPMLS 1090	

## RESULT 4

US-07-757-022B-52  
; Sequence 52, Application US/07757022B  
; Patent No. 6433142

; GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

COUNTRY: U.S.A.

; ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: 07/757,022B  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1363 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-52

Query Match.	99.7%;	Score 5561;	DB 4;	Length 1363;
Best local Similarity	95.5%;	Pred. No. 1.4e-169;		
Matches 1049;	Conservative 0;	Mismatches 0;	Indels 50;	Gaps 1
QY	1	MAMKTPILYLLLLSVFVIQOVSSOELSCKRCFESFERGECDCDAOCKKYDKCCPYE	60	
Db	1	MAMKTPILYLLLLSVFVIQOVSSOELSCKRCFESFERGECDCDAOCKKYDKCCPYE	60	
QY	61	SFCA-----EEHSVS 70		
Db	61	SFCAEVHNPTSPSSKAPPPSGASQIKSTTKSPPPNKKTKKVIESEIETEEHSVS	120	
QY	71	ENQESSSSSSSSSSSTIWKIKSKNSAANRELQKLVKDNKNKRTKKKPTKPPVVDE	130	
Db	121	ENQESSSSSSSSSSSIWKIKSKNSAANRELQKLVKDNKNKRTKKKPTKPPVVDE	180	
QY	131	AGSLDNGDFKVTPTDSTTQHNVKTSVPKITTAKPINRPSLSPNSDTSKETSILTVMKE	190	
Db	181	AGSLDNGDFKVTPTDSTTQHNVKTSVPKITTAKPINRPSLSPNSDTSKETSILTVMKE	240	
QY	191	TTVETKEITTNTKQTSIDGKEKTTSAKETOSIEKTSADKLAPTSKVLAKPTPKAETTTKG	250	
Db	241	TTVETKEITTNTKQTSIDGKEKTTSAKETOSIEKTSADKLAPTSKVLAKPTPKAETTTKG	300	
QY	251	PALTTTKEPTPTTPKEPASITTPKEPTTTIKSAPTTTKEPAPTTTKGAPTTTKEPAPTTT	310	
Db	301	PALTTTKEPTPTTPKEPASITTPKEPTTTIKSAPTTTKEPAPTTTKGAPTTTKEPAPTTT	360	
QY	311	KEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTKEPAPTTTPKKPAPTTTKEPAPTTTKEPPT	370	
Db	361	KEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTKEPAPTTTPKKPAPTTTKEPAPTTTKEPPT	420	
QY	371	TTTKEPAPTTTKEPAPTTTKEPAPTTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPPTTKE	430	
Db	421	TTTKEPAPTTTKEPAPTTTKEPAPTTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPPTTKE	480	
QY	431	PAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT	490	







QY 750 KPAPPTPEPPPTTSEVSTPTTKEPTTHKSPDESTPELSABPTPKALENSPKPCVPT 809  
Db 841 KPAPPTPEPPPTTSEVSTPTTKEPTTHKSPDESTPELSABPTPKALENSPKPCVPT 900  
QY 810 TKTPAAKPEMTTAKDKTTERDLRTPTPETTAAPKMTKETATTTKTTESKITATTQV 869  
Db 901 TKTPAAKPEMTTAKDKTTERDLRTPTPETTAAPKMTKETATTTKTTESKITATTQV 960  
QY 870 TSTTTQDTPPKITTLKTTTLAPKVTTTKTITTEIMNKPEBETAKPKDRATNSKATTPK 929  
Db 961 TSTTTQDTPPKITTLKTTTLAPKVTTTKTITTEIMNKPEBETAKPKDRATNSKATTPK 1020  
QY 930 POKPTKAPKPTSTKPKMVRVRKPTTPTPKMTSTMEELNPTSRIAEAMLQTTTRPN 989  
Db 1021 POKPTKAPKPTSTKPKMVRVRKPTTPTPKMTSTMEELNPTSRIAEAMLQTTTRPN 1080  
QY 990 QTPNSKLVEVNPKSEDAGGAEGETHMLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1049  
Db 1081 QTPNSKLVEVNPKSEDAGGAEGETHMLRPHVFMPEVTPDMDYLPRVNPQGIINPMLS 1140

RESULT 7

US-07-757-022B-62  
; Sequence 62, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1404 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
US-07-757-022B-62  
Query Match 99.5%; Score 5546.9; DB 4; Length 1404;  
Best Local Similarity 92.0%; Pred. No. 4.1e-169;  
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;  
QY 1 MAWKTLPIYLLILLLSVFIQVSSQ----- 25  
Db 1 MAWKTLPIYLLILLLSVFIQVSSQDLSSCAGCGEGYSRDATCNCYDNCQHYMECCPDF 60  
QY 26 -----ELSCGRCFESFERGRECDCAOQCKYDKCCPDYSEFCA----- 64  
Db 61 KRVTIAELSCGRCFESFERGRECDCAOQCKYDKCCPDYSEFCAEVHNPTSPSSKAP 120  
QY 65 -----BEHSVSENOESSSSSSSSSIW 89  
Db 121 PPSGASQITKSTTKRSPKPPNKKTKKVIESEBITTEHSVSENOESSSSSSSSSIW 180  
QY 90 KIKSSKNSAANRELOKKLVKONKKNRTKKKTPKPPVVDAGSLDNGDFKVTITDST 149  
Db 181 KIKSSKNSAANRELOKKLVKONKKNRTKKKTPKPPVVDAGSLDNGDFKVTITDST 240  
QY 150 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVANKETIVETKETTINKQSTSDG 209  
Db 241 TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVANKETIVETKETTINKQSTSDG 300  
QY 210 KECTSAKTSQTSIEKTSKDLAPTSKVLAKEPTKAEITTKGPAITTKPEPTTTPKEPAS 269  
Db 301 KECTSAKTSQTSIEKTSKDLAPTSKVLAKEPTKAEITTKGPAITTKPEPTTTPKEPAS 360  
QY 270 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 329  
Db 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTKPE 420  
QY 330 APITTKSAPTTKPEAPTTPKPEAPTTKPEAPTTKPEPTTTPKPEAPTTKPEAPTTPK 389  
Db 421 APITTKSAPTTKPEAPTTPKPEAPTTKPEAPTTKPEPTTTPKPEAPTTKPEAPTTPK 480  
QY 390 EPAPTAPKPEAPTTKPEAPTTKPEAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 449  
Db 481 EPAPTAPKPEAPTTKPEAPTTKPEAPTTTKPEPTTTPKPEAPTTTKSAPTTTKPEAPT 540  
QY 450 TTKSAPTTKPEPSPTTKPEAPTTKPEAPTTPKPEAPTTKPEAPTTKPEAPTTTKKP 509  
Db 541 TTKSAPTTKPEPSPTTKPEAPTTKPEAPTTPKPEAPTTPKPEAPTTKPEAPTTTKKP 600  
QY 510 APTAPKPEAPTTPKETAPTTPKKLTPTTPEKLAPTTPKPEAPTTPELAPTTPEEPTPTT 569  
Db 601 APTAPKPEAPTTPKETAPTTPKKLTPTTPEKLAPTTPKPEAPTTPELAPTTPEEPTPTT 660  
QY 570 PEEPAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTPKPEAPTTTPKGTAPTTTLKEP 629  
Db 661 PEEPAPTTPKAAAPNTKPEAPTTKPEAPTTKPEAPTTPKPEAPTTTPKGTAPTTTLKEP 720  
QY 630 APTTPKPKAPKELAPTTTKPEPTSTSDKAPTTPKGTAPTTKPEAPTTKPEAPTTPKG 689  
Db 721 APTTPKPKAPKELAPTTTKPEPTSTSDKAPTTPKGTAPTTKPEAPTTKPEAPTTPKG 780  
QY 690 TAPTTLKPEAPTTPKKAPKELAPTTTKGPTSTSDKAPTTPKPEAPTTPKETAPTTKPEAPTTPK 749  
Db 781 TAPTTLKPEAPTTPKKAPKELAPTTTKGPTSTSDKAPTTPKPEAPTTPKETAPTTKPEAPTTPK 840  
QY 750 KPAPTTPEPPPTTSEVSTPTTKEPTTHKSPDESTPELSABPTPKALENSPKPCVPT 809  
Db 841 KPAPTTPEPPPTTSEVSTPTTKEPTTHKSPDESTPELSABPTPKALENSPKPCVPT 900  
QY 810 TKTPAAKPEMTTAKDKTTERDLRTPTPETTAAPKMTKETATTTKTTESKITATTQV 869  
Db 901 TKTPAAKPEMTTAKDKTTERDLRTPTPETTAAPKMTKETATTTKTTESKITATTQV 960  
QY 870 TSTTTQDTPPKITTLKTTTLAPKVTTTKTITTEIMNKPEBETAKPKDRATNSKATTPK 929

Db 961 TSTTTQDTPFKITTLTKTTTLPAPKVVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 1020  
Qy 930 POKETKAPKPTSTKKPKMTMPVRKPKTTPTPKMTSTMPDELNPTSRIAEAMLOTTTRPN 989  
Db 1021 POKPTKAPKPTSTKKPKMTMPVRKPKTTPTPKMTSTMPDELNPTSRIAEAMLOTTTRPN 1080  
Qy 990 QTNSKLIVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1049  
Db 1081 QTNSKLIVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1140

RESULT 8  
US-09-298-970A-1  
; Sequence 1, Application US/09298970A  
; Patent No. 6743774  
; GENERAL INFORMATION:  
; APPLICANT: Jay, Gregory D.  
; TITLE OF INVENTION: TRIBONECTINS  
; FILE REFERENCE: 21486-026  
; CURRENT APPLICATION NUMBER: US/09/298,970A  
; CURRENT FILING DATE: 2001-06-19  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1404  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-298-970A-1

Query Match 99.5%; Score 5546.9; DB 4; Length 1404;  
Best Local Similarity 92.0%; Pred. No. 4.1e-169;  
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;  
Qy 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25  
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60  
Qy 26 -----ELSCCKGRCFESFERGREGCDCAOCKYKDYKCCPDYESFCA----- 64  
Db 61 KRVTAEELSCCKGRCFESFERGREGCDCAOCKYKDYKCCPDYESFCAEVHNFTSPSSKAP 120  
Qy 65 -----BEHSVSNQESSSSSSSSSSSTIW 89  
Db 121 PPASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVSNQESSSSSSSSSSSTIW 180  
Qy 90 KIKSSKNSAANRELQKKLVKDNKNKRTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 149  
Db 181 KIKSSKNSAANRELQKKLVKDNKNKRTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 240  
Qy 150 TQHNKYSTSPKITTAKDINPRPSLPNNSDTSKETSITVNETVETKETTITNKQSTDG 209  
Db 241 TQHNKYSTSPKITTAKDINPRPSLPNNSDTSKETSITVNETVETKETTITNKQSTDG 300  
Qy 210 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTKAEETTKGPAETTPKPEPTTPKBPAS 269  
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTKAEETTKGPAETTPKPEPTTPKBPAS 360  
Qy 270 TTPKEPTPTTIKSAPTTPKBPATTKSAPTTPKBPATTKSAPTTPKBPATTKSAPTTPKBPATTKKEP 329  
Db 361 TTPKEPTPTTIKSAPTTPKBPATTKSAPTTPKBPATTKSAPTTPKBPATTKSAPTTPKBPATTKKEP 420  
Qy 330 APITTKSAPTTPKBPATTKPKKAPPTPKBPATTPKPEPTTPKBPATTPKBPATTKBPATTPK 389  
Db 421 APITTKSAPTTPKBPATTPKBPATTPKBPATTPKPEPTTPKBPATTPKBPATTKBPATTPK 480  
Qy 390 EPAPAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTKBPATTPK 449  
Db 481 EPAPAPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTKBPATTPK 540  
Qy 450 TTKSAPTTPKBPSTTTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTKBP 509  
Db 541 TTKSAPTTPKBPSTTTTKBPATTPKBPATTPKBPATTPKBPATTPKBPATTPKBPATTKBP 600

Qy 510 APTAPKEPAPTTPEKETAPTTPKKLTPTTPKELAPTTPEKAPATTPPEELAPTTPEEPPTPT 569  
Db 601 APTAPKEPAPTTPEKETAPTTPKKLTPTTPKELAPTTPEKAPATTPPEELAPTTPEEPPTPT 660  
Qy 570 PESPAPTTPKAAAPNTPKBPAPTTPEKAPATTPPEKAPATTPPKETAPTTPKGTAPTTLKEP 629  
Db 661 PESPAPTTPKAAAPNTPKBPAPTTPEKAPATTPPEKAPATTPPKETAPTTPKGTAPTTLKEP 720  
Qy 630 APTTPKKAPKELAPTTTKEPTSTTSKAPATTPKGTAPTTPEKAPATTPPEKAPATTPPKG 689  
Db 721 APTTPKKAPKELAPTTTKEPTSTTSKAPATTPKGTAPTTPEKAPATTPPEKAPATTPPKG 780  
Qy 690 TAPTTLKEPAPTTPKAPKELAPTTTKEPTSTTSKAPATTPPKETAPTTPEKAPATTPPK 749  
Db 781 TAPTTLKEPAPTTPKAPKELAPTTTKEPTSTTSKAPATTPPKETAPTTPEKAPATTPPK 840  
Qy 750 KPAPTTPEPTPTTSEVSTPTTTTKEPTTIHKSDESTPELSASPTPKALNSPKPEVPT 809  
Db 841 KPAPTTPEPTPTTSEVSTPTTTTKEPTTIHKSDESTPELSASPTPKALNSPKPEVPT 900  
Qy 810 TKTPAATKPEMTTAKDKTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQV 869  
Db 901 TKTPAATKPEMTTAKDKTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQV 960  
Qy 870 TSTTTQDTPFKITTLTKTTTLPAPKVVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 929  
Db 961 TSTTTQDTPFKITTLTKTTTLPAPKVVTTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK 1020  
Qy 930 POKPTKAPKPTSTKKPKMTMPVRKPKTTPTPKMTSTMPDELNPTSRIAEAMLOTTTRPN 989  
Db 1021 POKPTKAPKPTSTKKPKMTMPVRKPKTTPTPKMTSTMPDELNPTSRIAEAMLOTTTRPN 1080  
Qy 990 QTNSKLIVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1049  
Db 1081 QTNSKLIVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMOYLPRVFNQGIINPMLS 1140

RESULT 9  
US-10-164-595-78  
; Sequence 78, Application US/10164595  
; Patent No. 6657054  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies, Inc  
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
; FILE REFERENCE: IU 103 R1  
; CURRENT APPLICATION NUMBER: US/10/164,595  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 78  
; LENGTH: 1404  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-595-78

Query Match 99.1%; Score 5523.9; DB 4; Length 1404;  
Best Local Similarity 91.8%; Pred. No. 2.2e-168;  
Matches 1046; Conservative 0; Mismatches 3; Indels 91; Gaps 2;  
Qy 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25  
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHYMECCPDF 60  
Qy 26 -----ELSCCKGRCFESFERGREGCDCAOCKYKDYKCCPDYESFCA----- 64  
Db 61 KRVTAEELSCCKGRCFESFERGREGCDCAOCKYKDYKCCPDYESFCAEVHNFTSPSSKAP 120  
Qy 65 -----BEHSVSNQESSSSSSSSSSSTIW 89  
Db 121 PPASQTIKSTTKRSPKPNKKTKKVIIESEITEHSVSNQESSSSSSSSSSSTIW 180  
Qy 90 KIKSSKNSAANRELQKKLVKDNKNKRTKKPKPPVVDVDEAGSLDNGDFKVTTPDTST 149



[illegible]

## RESULT 11

RESOLUT  
US-07-757-0228-46

03-07-757-022B-48  
: Sequence 46. Application US/07757022B

; Patent No. 6433142

FACEID NO. 0433142  
; GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

1. TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

```

: MEDIUM TYPE: Floppy disk
: COME FROM READABLE FORM,

```

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-

SOFTWARE: Patent In Release #1.0. Version #1.25

SOFTWARE: FACEGATE RELEASE #1.0; V  
: CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757.022B

```

/ FILING DATE: 19910910
/ CLASSIFICATION: 530
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/643,502
/ FILING DATE: 18-JAN-1991
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/546,114
/ FILING DATE: 29-JUN-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/457,196
/ FILING DATE: 29-DEC-1989
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/390,901
/ FILING DATE: 08-AUG-1989
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Cseir, Luann
/ REGISTRATION NUMBER: 31,822
/ REFERENCE/DOCKET NUMBER: GI 5190
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)876-1170
/ TELEFAX: (617)876-5851
/ INFORMATION FOR SEQ ID NO: 46:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1320 amino acids
/ TYPE: AMINO ACID
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-07-757-022B-46

```

Query Match	96.7%	Score 5393.3	DB 4	Length 1320	
Best Local Similarity	96.1%	Pred. No. 3e-164			
Matches 1015	Conservative	8	Mismatches 26	Indels 7	Gaps 1
Qy	1	MAWKTLPIVILLLLSVFVIQQVSSQBSLCKGRCFESFERGECDDAQCKYDKCCPDYE	60		
Db	1	MAWKTLPIVILLLLSVFVIQQVSSQBSLCKGRCFESFERGECDDAQCKYDKCCPDYE	60		
Qy	61	SFCAEHSVSENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKKL-----KYKDNK	113		
Db	61	SFCAEVHNTSPSSKKAPPGASQIKSTTKRSPKPPNKTKKVIESEEITEVKDNK	120		
Qy	114	KXRTKKKPTPKPVVDGAGSLNDGDFKVTTPDTSITQHNVKSTSPKIIITAKPINRPSL	173		
Db	121	KXRTKKKPTPKPVVDGAGSLNDGDFKVTTPDTSITQHNVKSTSPKIIITAKPINRPSL	180		
Qy	174	PNNSDTSKETSILVNKETTIVETKETTNNKQTSIDGKEKTTSAKETQSIEKTSAKDLAPT	233		
Db	181	PNNSDTSKETSILVNKETTIVETKETTNNKQTSIDGKEKTTSAKETQSIEKTSAKDLAPT	240		
Qy	234	SKVLAKPTPKAETTTKGPALTTPKETPTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPT	293		
Db	241	SKVLAKPTPKAETTTKGPALTTPKETPTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPT	300		
Qy	294	TTKSAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTTPKKPA	353		
Db	301	TTKSAPTTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTTSAPTTTPKEPAPTTPKKPA	360		
Qy	354	PTTPKEPAPTTTPKEPPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTAPKKEPAPTTTPKEPAPTTTPK	413		
Db	361	PTTPKEPAPTTTPKEPPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTAPKKEPAPTTTPKEPAPTTTPK	420		
Qy	414	EPAPTTTKESPPTTPKEPAPTTTKSAPTTTKGPAPTTTKSAPTTTPKEPSPPTTKKEPAPTT	473		
Db	421	EPAPTTTKESPPTTPKEPAPTTTKSAPTTTKGPAPTTTKSAPTTTPKEPSPPTTKKEPAPTT	480		
Qy	474	PKEPAPTTPKKAPPTTPKEPAPTTTPKEPAPTTTKKPAAPTAPKEPAPTTTKEPAPTTTPKKL	533		
Db	481	PKEPAPTTPKKAPPTTPKEPAPTTTPKEPAPTTTKKPAAPTAPKEPAPTTTKEPAPTTTPKKL	540		
Qy	534	TTPTPKLAPTTTPKEPAPTTTPBELAPTTTPEEPTTPTTPEEAPTTTPKAAAPNTPKBPAPT	593		
Db	541	TTPTPKLAPTTTPKEPAPTTTPBELAPTTTPEEPTTPTTPEEAPTTTPKAAAPNTPKBPAPT	600		



QY 834 RTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTTQDTPPKITLTKTTTLAPK 893  
DB 841 RTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTTQDTPPKITLTKTTTLAPK 900  
QY 894 VTTTTKTITTTTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKXSTSKKPKTMPRVR 953  
DB 901 VTTTTKTITTTTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKXSTSKKPKTMPRVR 960  
QY 954 KPXTTTPRKNWSTMPBLNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGAGSET 1013  
DB 961 KPXTTTPRKNWSTMPBLNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGAGSET 1020  
QY 1014 PHMLLRPHVFMPEVTPDMVDYLPRVFNQGIINPMLS 1049  
DB 1021 PHMLLRPHVFMPEVTPDMVDYLPRVFNQGIINPMLS 1056  
RESULT 13  
US-10-164-595-58  
; Sequence 58, Application US/10164595  
; Patent No. 6657054  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies, Inc  
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
; FILE REFERENCE: IU 103 R1  
; CURRENT APPLICATION NUMBER: US/10/164,595  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 58  
; LENGTH: 1320  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-595-58  
Query Match 96.6%; Score 5384.3; DB 4; Length 1320;  
Best Local Similarity 95.9%; Pred. No. 5.8e-164;  
Matches 1013; Conservative 8; Mismatches 28; Indels 7; Gaps 1;  
QY 1 MAWKTLPIYLLLSVFIQVSSQELSCKGRCPESFERGRCDCDAQCKYDKCCPDYE 60  
DB 1 MAWKTLPIYLLLSVFIQVSSQELSCKGRCPESFERGRCDCDAQCKYDKCCPDYE 60  
QY 61 SFCAEHSVSENQSSSSSSSSSTTWIKSSKNSANRELQKL-----KVQDNK 113  
DB 61 SFCAEHNPTSPSSKKAPPPSGASQTIKSTTKRSPKPPNKKTKKVIESEITEVDKNK 120  
QY 114 KNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDTSSTQHNVSTSPKITTAKPINRPSL 173  
DB 121 KNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDTSSTQHNVSTSPKITTAKPINRPSL 180  
QY 174 PPSDTSKERSLVNKEETTVEKTTTNTQTSQDGEKTTSAKETQSIEKTSADLAPT 233  
DB 181 PPSDTSKERSLVNKEETTVEKTTTNTQTSQDGEKTTSAKETQSIEKTSADLAPT 240  
QY 234 SKVLAKPTPKAETTKGPALTTPKEPTTPKPEASTTPKEPTTTIKSAPTTPKEAPT 293  
DB 241 SKVLAKPTPKAETTKGPALTTPKEPTTPKPEASTTPKEPTTTIKSAPTTPKEAPT 300  
QY 294 TTKSAPTTPKEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPKPA 353  
DB 301 TTKSAPTTPKEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPKPA 360  
QY 354 PTPKPEAPTTPKEPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 413  
DB 361 PTPKPEAPTTPKEPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 420  
QY 414 EPAPTTHKPSPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 473  
DB 421 EPAPTTHKPSPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPT 480  
QY 474 PKEAPTTPPKKPAETTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPKL 533

DB 481 PKEAPTTPPKKPAETTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPKL 540  
QY 534 TPTTPEKLAETTPKPEAPTTPPELAPTTPEEPPTTPPEEPAPTTPKAAANPTPKPEAPT 593  
DB 541 TPTTPEKLAETTPKPEAPTTPPELAPTTPEEPPTTPPEEPAPTTPKAAANPTPKPEAPT 600  
QY 594 PKEAPTTPKPEAPTTPKETAAPTTPKGTAPTTLKEAPATTPKPKAPKELAPTTTKEPTST 653  
DB 601 PKEAPTTPKPEAPTTPKETAAPTTPKGTAPTTLKEAPATTPKPKAPKELAPTTTKEPTST 660  
QY 654 TSOKPAPTTPKGTAPTTPKPEAPTTPKPEAPTTPKGTAPTTLKEAPATTPKPKAPKELAP 713  
DB 661 TCDKAPTTPKGTAPTTPKPEAPTTPKPEAPTTPKGTAPTTLKEAPATTPKPKAPKELAP 720  
QY 714 TTTKGTSTTSOKPAPTTPKETAAPTTPKPEAPTTPKPEAPTTPKETAAPTTPPEPTTSEVSTPTTK 773  
DB 721 TTTKGTSTTSOKPAPTTPKETAAPTTPKPEAPTTPKPEAPTTPKETAAPTTPPEPTTSEVSTPTTK 780  
QY 774 EPTTIHKSPDESTPELSAETPKALENSPKPEPGVPTTKTAAATKPEMTTTAKDKTTBRDL 833  
DB 781 EPTTIHKSPDESTPELSAETPKALENSPKPEPGVPTTKTAAATKPEMTTTAKDKTTBRDL 840  
QY 834 RTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTTQDTPPKITLTKTTTLAPK 893  
DB 841 RTTPEITTAAPKMTKETATTTTEKTESKITATTTQVSTTTTQDTPPKITLTKTTTLAPK 900  
QY 894 VTTTTKTITTTTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKXSTSKKPKTMPRVR 953  
DB 901 VTTTTKTITTTTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKXSTSKKPKTMPRVR 960  
QY 954 KPXTTTPRKNWSTMPBLNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGAGSET 1013  
DB 961 KPXTTTPRKNWSTMPBLNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGAGSET 1020  
QY 1014 PHMLLRPHVFMPEVTPDMVDYLPRVFNQGIINPMLS 1049  
DB 1021 PHMLLRPHVFMPEVTPDMVDYLPRVFNQGIINPMLS 1056  
RESULT 14  
US-07-757-022B-40  
; Sequence 40, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/07/757,022B  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1361 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-40

Query Match          36.5%; Score 5379.2; DB 4; Length 1361;
Best Local Similarity 92.5%; Pred. No. 8.7e-164;
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;

QY 1 MANKTPIYLLLLSVFVIOQVSSQ-----25
DB 1 MANKTPIYLLLLSVFVIOQVSSQDLSSCAGCGEGYSRDATAICNDYCNQHMECCPDF 60

QY 26 -----ELSCGRCFESFERGRCDCDAQCKYDKCCPDYEFSCAEHSHSVSENQESSSS 79
DB 61 KRVCTAELSCGRCFESFERGRCDCDAQCKYDKCCPDYEFSCAEVHNFTSPSSKKAP 120

QY 80 SSSSSSTIWKISKNSAANRELQKL-----KVONKQKRTKKKPTPKPPVDEAG 132
DB 121 PPSGASQITKTRSPKPPKTKKVISEIEVEKDNKQKRTKKKPTPKPPVDEAG 180

QY 133 SGLDNGDFKVTTPDSTQHNKYSTSPKITTAKPINRPSLPPNSDTSKETSIVNKETT 192
DB 181 SGLDNGDFKVTTPDSTQHNKYSTSPKITTAKPINRPSLPPNSDTSKETSIVNKETT 240

QY 193 VETKETITNNKQSTDKGKXITSAKETQSIKTSKADLAFTSKVLAKPTKAEITTKGPA 252
DB 241 VETKETITNNKQSTDKGKXITSAKETQSIKTSKADLAFTSKVLAKPTKAEITTKGPA 300

QY 253 LITPKETPTTPKEPASTTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE 312
DB 301 LATPKETPTTPKEPASTTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKE 360

QY 313 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTTPKPEAPTTTPKPEAPTTTKEPTPT 372
DB 361 PAPTTPKEAPTTTKPEAPTTTKSAPTTKPEAPTTTPKPEAPTTTPKPEAPTTTKEPTPT 420

QY 373 PKPEAPTTKEPAPTTPKEPAPTAPKPEAPTTTPKPEAPTTTPKPEAPTTTKFSPPTTKEPA 432
DB 421 PKPEAPTTKEPAPTTPKPEAPTTAPKPEAPTTTPKPEAPTTTPKPEAPTTTKFSPPTTKEPA 480

QY 433 PTTTKSAPTTTKPEAPTTTKSAPTTTPKEPSPTTKEPAPTTPKPEAPTTTPKPEAPTTTKE 492
DB 481 PTTTKSAPTTTKPEAPTTTKSAPTTTPKEPSPTTKEPAPTTPKPEAPTTTPKPEAPTTTKE 540

QY 493 PAPTTPKEAPTTTKPEAPTTAPKPEAPTTTPKETAAPTTPKLTPTTPEKLAAPTTPKEAPTT 552
DB 541 PAPTTPKEAPTTTKPEAPTTAPKPEAPTTTPKETAAPTTPKLTPTTPEKLAAPTTPKEAPTT 600

QY 553 TPBELAPTTPPEEPTPTTPEBPAPTTPKAAANPTPKPEAPTTTPKPEAPTTTPKPEAPTTTKE 612
DB 601 TPBELAPTTPPEEPTPTTPEBPAPTTPKAAANPTPKPEAPTTTPKPEAPTTTPKPEAPTTTKE 660

QY 613 TAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTPKETSITSDKPAPTTPKGTAPTTPK 672
DB 661 TAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTPKETSITSDKPAPTTPKGTAPTTPK 720

```

## RESULT 15

US-07-757-022B-74

; Sequence 74, Application US/07/757022B

; Patent No. 6433142

; GENERAL INFORMATION:

; APPLICANT: Gesner, Thomas G.

; APPLICANT: Clark, Stephen C.

; APPLICANT: Turner, Katherine

; APPLICANT: Hewick, Rodney M.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/757,022B

; FILING DATE: 19910910

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/643,502

; FILING DATE: 18-JAN-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/546,114

; FILING DATE: 29-JUN-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/457,196

; FILING DATE: 29-DEC-1989

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/390,901

; FILING DATE: 08-AUG-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: Cserr, Luann

; REGISTRATION NUMBER: 31,822

; REFERENCE/DOCKET NUMBER: GI 5190

; TELECOMMUNICATION INFORMATION:





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 106.723 Seconds

(without alignments)  
3171.696 Million cell updates/sec

Title: SEQ1-D

Perfect score: 5576

Sequence: 1 MAWKTLPIYILLLSVFVIQ.....DMDYLPVFNQGIINPMLS 1049

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1356558 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5576	100.0	1049	13	US-10-124-557-58
2	5576	100.0	1313	13	US-10-124-557-142
3	5561.9	99.7	1354	13	US-10-124-557-48
4	5561	99.7	1363	13	US-10-124-557-52
5	5546.9	99.5	1140	13	US-10-124-557-104
6	5546.9	99.5	1404	9	US-09-802-207-30
7	5546.9	99.5	1404	11	US-09-897-188-1
8	5546.9	99.5	1404	13	US-10-124-557-2
9	5546.9	99.5	1404	13	US-10-124-557-62
10	5447.9	97.7	1314	13	US-10-124-557-50
11	5393.3	96.7	1320	13	US-10-124-557-46
12	5393.3	96.7	1320	13	US-10-124-557-60
13	5379.2	96.5	1361	13	US-10-124-557-40
14	5358.7	96.1	1038	13	US-10-124-557-74

15 5358.7 96.1 1270 13 US-10-124-557-44  
16 5344.6 95.9 1311 13 US-10-124-557-42  
17 5235.9 93.9 1022 13 US-10-124-557-84  
18 5011 89.9 941 13 US-10-124-557-14  
19 2182.7 39.1 792 9 US-09-802-207-27  
20 1409.9 25.3 538 14 US-10-038-694-3  
21 1346.8 24.2 5179 9 US-09-922-217-1068  
22 1346.8 24.2 5179 9 US-09-833-263-1068  
23 1346.8 24.2 5179 13 US-10-025-380-1068  
24 1346.8 24.2 5179 16 US-10-734-564-121  
25 1174.4 21.1 292 16 US-10-468-910-4  
26 951.1 17.1 1460 14 US-10-295-027-428  
27 931.8 16.7 1325 9 US-09-864-761-35612  
28 918 16.5 1367 9 US-09-801-368-108  
29 860.6 15.4 3507 14 US-10-369-493-5784  
30 857.5 15.4 6642 14 US-10-369-493-5013  
31 855.3 15.3 2090 16 US-10-408-765A-2318  
32 834.9 15.0 3256 16 US-10-408-765A-174  
33 834.9 15.0 3256 16 US-10-701-490-9  
34 832.7 14.9 697 15 US-10-425-114-41545  
35 828.7 14.9 19723 15 US-10-084-846A-5  
36 827.1 14.8 3256 9 US-09-919-172-98  
37 827.1 14.8 3256 10 US-09-919-039-21  
38 822.8 14.8 1151 10 US-09-825-751A-79  
39 822 14.7 2397 15 US-10-282-122A-71232  
40 817.1 14.7 22152 16 US-10-715-066-5  
41 816.4 14.6 5935 14 US-10-243-243A-8  
42 816 14.6 5877 14 US-10-142-515-11  
43 809.6 14.5 2971 14 US-10-146-473-50  
44 799.7 14.3 1255 10 US-09-996-069-10  
45 799.7 14.3 1255 14 US-10-171-311-158

#### ALIGNMENTS

#### RESULT 1

US-10-124-557-58

; Sequence 58, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

APPLICANT: Turner, Katherine

Clark, Stephen C.

Jacobs, Kenneth

Hewick, Rodney M.

Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION NUMBER: US/10124,557

FILING DATE: 16-Apr-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1049 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 58:
US-10-124-557-58

Query Match 100.0%; Score 5576; DB 13; Length 1049;
Best Local Similarity 100.0%; Pred. No. 1.1e-139;
Matches 1049; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKLPYVLLVSVFVTVQVSSQELSKGRCFCFSEFGRGRCDCDAQCKKYDKCCPDYE 60
Db 1 MAWKLPYVLLVSVFVTVQVSSQELSKGRCFCFSEFGRGRCDCDAQCKKYDKCCPDYE 60
QY 61 SFCAREHSVSENOESSSSSSSSSSSTIWKIKSSKNSAANRELOKKLVKDKNKNRTKKX 120
Db 61 SFCAREHSVSENOESSSSSSSSSSSTIWKIKSSKNSAANRELOKKLVKDKNKNRTKKX 120
QY 121 PTPKPPVVDVAGSGLDNGDFKVTTPDTSTTQHNKSVSTPKITPAKPINRPSLPPNSDTS 180
Db 121 PTPKPPVVDVAGSGLDNGDFKVTTPDTSTTQHNKSVSTPKITPAKPINRPSLPPNSDTS 180
QY 181 KETSLTVNKETTVKETTNNKQSTGDKKETSIAKETQSIBKTSKADLAPTSKVLAKP 240
Db 181 KETSLTVNKETTVKETTNNKQSTGDKKETSIAKETQSIBKTSKADLAPTSKVLAKP 240
QY 241 TPKAETTTKGPAUTTPKEPTPTTPKEPASTTPKEPTPTTKSAPTTKKEPAPTTPKSAPT 300
Db 241 TPKAETTTKGPAUTTPKEPTPTTPKEPASTTPKEPTPTTKSAPTTKKEPAPTTPKSAPT 300
QY 301 TPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKKEPAPTTPKKEPAPTTPKEP 360
Db 301 TPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKKEPAPTTPKKEPAPTTPKEP 360
QY 361 APPTKEPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKKEPAPTTPKEP 420
Db 361 APPTKEPPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKKEPAPTTPKEP 420
QY 421 KEPSPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 480
Db 421 KEPSPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 480
QY 481 TPKKPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 540
Db 481 TPKKPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 540
QY 541 LAPTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 600
Db 541 LAPTTKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 600
QY 601 TPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 660
Db 601 TPKEPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 660
QY 661 TTPKGAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 720
Db 661 TTPKGAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 720
QY 721 STTSKPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 780
Db 721 STTSKPAPTTPKEPAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPTTPKSAPT 780
QY 781 SPDESTPELSAETPKALENSPKPGVPTTKTPAATKPEMTTTAKDKXTERDLRTTPTT 840

```

Query Match		100.0%;	Score 5576;	DB 13;	Length 1313;				
Best Local Similarity		100.0%;	Pred. No. 1.4e-139;						
Matches 1049;		Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	MAWKTLPIYLLILLISVFVIOQVSSQELSCKGRCFESPERGREGCDCAQCKKYDKCCPDYE	60						
DB	1	MAWKTLPIYLLILLISVFVIOQVSSQELSCKGRCFESPERGREGCDCAQCKKYDKCCPDYE	60						
QY	61	SFCAEHSVSENOESSSSSSSSSSSTIWKIKSKNSAANRELOKLVKNDKNKRTKKK	120						
DB	61	SFCAEHSVSENOESSSSSSSSSSSTIWKIKSKNSAANRELOKLVKNDKNKRTKKK	120						
QY	121	PTKPPVVDAGSLDNGDFKVTTPDSTTTQHNKVSTSPKITTAKPINRPSPSPNSDTS	180						
DB	121	PTKPPVVDAGSLDNGDFKVTTPDSTTTQHNKVSTSPKITTAKPINRPSPSPNSDTS	180						
QY	181	KETSLVNKETTVETKETTNNKQSTSDGKEKTSKETSIAKOLAPTSKVLAKP	240						
DB	181	KETSLVNKETTVETKETTNNKQSTSDGKEKTSKETSIAKOLAPTSKVLAKP	240						
QY	241	TPKAEITTKGPALTTKPEPTTTPKBPASTTPKPEPTTTIKSAPTTPKBPATTTKSAPT	300						
DB	241	TPKAEITTKGPALTTKPEPTTTPKBPASTTPKPEPTTTIKSAPTTPKBPATTTKSAPT	300						
QY	301	TPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	360						
DB	301	TPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	360						
QY	361	APTTPKEPTTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT	420						
DB	361	APTTPKEPTTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT	420						
QY	421	KEPSPPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	480						
DB	421	KEPSPPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	480						
QY	481	TPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	540						
DB	481	TPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	540						
QY	541	LAPTTPEKAPPTPELAPTTPEPTTTPKBPATTTKBPATTTKBPATTTKBPATTT	600						
DB	541	LAPTTPEKAPPTPELAPTTPEPTTTPKBPATTTKBPATTTKBPATTTKBPATTT	600						
QY	601	TPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	660						
DB	601	TPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	660						
QY	661	TPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	720						
DB	661	TPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP	720						
QY	721	STTSKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT	780						
DB	721	STTSKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT	780						
QY	781	SPDESTPELSABPTPKALENSKPEGVPTTKPAATKPEMTTTAKDITERRDLRTPT	840						
DB	781	SPDESTPELSABPTPKALENSKPEGVPTTKPAATKPEMTTTAKDITERRDLRTPT	840						
QY	841	TAAPKMTKETAATTTKTTESKITATTQVTSSTTTQDTPFKITLLKTTTLAPKVTTTKT	900						
DB	841	TAAPKMTKETAATTTKTTESKITATTQVTSSTTTQDTPFKITLLKTTTLAPKVTTTKT	900						
QY	901	ITTTIIMNKPESTAKPKORATNSKATTPKQPTTAPKPTSTTKPKTMVVRKPKTTPT	960						
DB	901	ITTTIIMNKPESTAKPKORATNSKATTPKQPTTAPKPTSTTKPKTMVVRKPKTTPT	960						
QY	961	PRKMTSTWPELNPSTRIAEAMIQTTTRPNQTPNSKLVEVNPKSEDAGGEGTPEMLLRP	1020						
DB	961	PRKMTSTWPELNPSTRIAEAMIQTTTRPNQTPNSKLVEVNPKSEDAGGEGTPEMLLRP	1020						
QY	1021	HVFMEVTPDMDYLPVFNQGIINPMLS	1049						

DB	1021	HVFMEVTPDMDYLPVFNQGIINPMLS	1049						
RESULT 3									
US-10-124-557-48									
; Sequence 48, Application US/10124557									
; Publication No. US20020137894A1									
; GENERAL INFORMATION:									
; APPLICANT: Turner, Katherine C.									
; Clark, Stephen C.									
; Jacobs, Kenneth									
; Hewick, Rodney M.									
; Gesner, Thomas G.									
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors									
; NUMBER OF SEQUENCES: 143									
; CORRESPONDENCE ADDRESS:									
; ADDRESSEE: Genetics Institute, Inc.									
; STREET: 87 CambridgePark Drive									
; CITY: Cambridge									
; STATE: Massachusetts									
; COUNTRY: U.S.A.									
; ZIP: 02140									
; COMPUTER READABLE FORM:									
; MEDIUM TYPE: Floppy disk									
; COMPUTER: IBM PC compatible									
; OPERATING SYSTEM: PC-DOS/MS-DOS									
; SOFTWARE: Patent In Release #1.0, Version #1.25									
; CURRENT APPLICATION DATA:									
; APPLICATION NUMBER: US/10/124,557									
; FILING DATE: 16-Apr-2002									
; CLASSIFICATION: <Unknown>									
; PRIOR APPLICATION DATA:									
; APPLICATION NUMBER: US 07/643,502									
; FILING DATE: 18-JAN-1991									
; APPLICATION NUMBER: US 07/546,114									
; FILING DATE: 29-JUN-1990									
; APPLICATION NUMBER: US 07/457,196									
; FILING DATE: 29-DEC-1989									
; APPLICATION NUMBER: US 07/390,901									
; FILING DATE: 08-AUG-1989									
; ATTORNEY/AGENT INFORMATION:									
; NAME: Cserr, Luann									
; REGISTRATION NUMBER: 31,822									
; REFERENCE/DOCKET NUMBER: GI 5190									
; TELECOMMUNICATION INFORMATION:									
; TELEPHONE: (617) 876-1170									
; TELEFAX: (617) 876-5851									
; INFORMATION FOR SEQ ID NO: 48:									
; SEQUENCE CHARACTERISTICS:									
; LENGTH: 1354 amino acids									
; TYPE: amino acid									
; TOPOLOGY: linear									
; MOLECULE TYPE: protein									
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:									
US-10-124-557-48									
Query Match 99.7%; Score 5561.9; DB 13; Length 1354;									
Best Local Similarity 96.2%; Pred. No. 3.4e-139;									
Matches 1049; Conservative 0; Mismatches 0; Indels 41; Gaps 1;									
QY	1	MAWKTLPIYLLILLISVFVIOQVSSQ-----	25						
DB	1	MAWKTLPIYLLILLISVFVIOQVSSQ-----	25						
QY	26	-----ELSCKRCFCSPERGREGCDCAQCKKYDKCCPDYEFCAEHSVSENOESSSS	79						
DB	61	KRVCTAELSCKRCFCSPERGREGCDCAQCKKYDKCCPDYEFCAEHSVSENOESSSS	120						
QY	80	SSSSSSSTIWKIKSKNSAANRELOKLVKNDKNKRTKKPTPKPPVVDAGSLDNGD	139						
DB	121	SSSSSSSTIWKIKSKNSAANRELOKLVKNDKNKRTKKPTPKPPVVDAGSLDNGD	180						

QY 140 PKVTPDSTTQHNKVSFSPKLTAKPINPRSLPNSDTSKETSITVNNKETTVEKETT 199  
Db 181 FKVTPDSTTQHNKVSFSPKLTAKPINPRSLPNSDTSKETSITVNNKETTVEKETT 240  
QY 200 TTNKQTSIDGKEKTTSAKETSQISAKLAPTSKVLAKPAPKAEITTKGPAITTPKEP 259  
Db 241 TTNKQTSIDGKEKTTSAKETSQISAKLAPTSKVLAKPAPKAEITTKGPAITTPKEP 300  
QY 260 TPTTPKEPASTTPKBPPTTKSAPTTKSEAPTTKSAPTTPKSEAPTTKSEAPTTK 319  
Db 301 TPTTPKEPASTTPKBPPTTKSAPTTKSEAPTTKSAPTTPKSEAPTTKSEAPTTK 360  
QY 320 EPAPTTTKPAPTTKSAPTTPKSEAPTTKSAPTTPKSEAPTTKSEAPTTKSEAPTT 379  
Db 361 EPAPTTTKPAPTTKSAPTTPKSEAPTTKSAPTTPKSEAPTTKSEAPTTKSEAPTT 420  
QY 380 TKEPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKSA 439  
Db 421 TKEPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKSA 480  
QY 440 PTTTPKEPAPTTKSAPTTPKSEAPTTKSEAPTTKSEAPTTKSEAPTTKSEAPTT 499  
Db 481 PTTTPKEPAPTTKSAPTTPKSEAPTTKSEAPTTKSEAPTTKSEAPTTKSEAPTT 540  
QY 500 EPAPTTTKPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 559  
Db 541 EPAPTTTKPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 600  
QY 560 TTPPEPTTPPTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTT 619  
Db 601 TTPPEPTTPPTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTT 660  
QY 620 GTAPTTLKBPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 679  
Db 661 GTAPTTLKBPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 720  
QY 680 KBPAPTTKGPAPTTLKBPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 739  
Db 721 KBPAPTTKGPAPTTLKBPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 780  
QY 740 KPEPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 799  
Db 781 KPEPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAPTTKAPAP 840  
QY 800 NSPKFQVPTTKPAATKPEMTTAKDKITTERDLRTTPTTAAKPMKETAATTTKETE 859  
Db 841 NSPKFQVPTTKPAATKPEMTTAKDKITTERDLRTTPTTAAKPMKETAATTTKETE 900  
QY 860 SKITATTTQVSTTQDTPFKITTLKTTTLPKVTITTKITITTEIMNKPEETAKPKDR 919  
Db 901 SKITATTTQVSTTQDTPFKITTLKTTTLPKVTITTKITITTEIMNKPEETAKPKDR 960  
QY 920 ATNSKATTPKPKQPTKAPKPTSTKPKTMKPRVRKPKTTPTRKMTSTMPELNPTSAIE 979  
Db 961 ATNSKATTPKPKQPTKAPKPTSTKPKTMKPRVRKPKTTPTRKMTSTMPELNPTSAIE 1020  
QY 980 AMLQTTTRPNOTNSKLVEYNPKSEDAGGAGETPHMLLRPHVPMPEVTPMDYLPVNP 1039  
Db 1021 AMLQTTTRPNOTNSKLVEYNPKSEDAGGAGETPHMLLRPHVPMPEVTPMDYLPVNP 1080  
QY 1040 QGIIINPMLS 1049  
Db 1081 QGIIINPMLS 1090

RESULT 4  
US-10-124-557-52  
; Sequence 52, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth

Hewick, Rodney M.  
Gesner, Thomas G.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1363 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 52:  
US-10-124-557-52  
Query Match 99.7%; Score 5561; DB 13; Length 1363;  
Best Local Similarity 95.5%; Pred. No. 3.6e-139;  
Matches 1049; Conservative 0; Mismatches 0; Indels 50; Gaps 1;  
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQBELSKGRCFESFERGECDCDAQCKKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQBELSKGRCFESFERGECDCDAQCKKYDKCCPDYE 60  
QY 61 SFCA-----EHSVS 70  
Db 61 SFCAEHNPTSPSPSKKAPPSPGASQITKTTKRSPKPNKKTKKVIIESEITEHSVS 120  
QY 71 ENQESSSSSSSSSSSTIWKIKSSKNSAANRELOKLVKDNKNRTKKKTPKPPVVD 130  
Db 121 ENQESSSSSSSSSSSTIWKIKSSKNSAANRELOKLVKDNKNRTKKKTPKPPVVD 180  
QY 131 AGSGLONGDFKVTTPDTSTTQHNKVSFSPKLTAKPINPRSLPNSDTSKETSITVNNKE 190  
Db 181 AGSGLONGDFKVTTPDTSTTQHNKVSFSPKLTAKPINPRSLPNSDTSKETSITVNNKE 240  
QY 191 TTVEKETTNTKQSTDGHEKTTSAKETOSIEKTSKADLAPTSKVLAKPAPKAEITTKG 250  
Db 241 TTVEKETTNTKQSTDGHEKTTSAKETOSIEKTSKADLAPTSKVLAKPAPKAEITTKG 300  
QY 251 PALTTTPKEPTTPKBPASTTPKEPTTTIKSAPTTKSEAPTTKSEAPTTKSEAPTT 310  
Db 301 PALTTTPKEPTTPKBPASTTPKEPTTTIKSAPTTKSEAPTTKSEAPTTKSEAPTT 360





```

RESULT 7
US-09-897-188-1
; Sequence 1, Application US/09897188
; Publication No. US20040072741A1
; GENERAL INFORMATION:
; APPLICANT: Jay, Gregory D.
; TITLE OF INVENTION: Tribonectin Polypeptides and Uses Thereof
; FILE REFERENCE: 21486-026 CIP2
; CURRENT APPLICATION NUMBER: US/09/897,188
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 09/298,970
; PRIOR FILING DATE: 1999-04-23
; PRIOR APPLICATION NUMBER: 09/556,246
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-188-1

Query Match          99.5%; Score 5546.9; DB 11; Length 1404;
Best Local Similarity 92.0%; Pred. No. 8.7e-139;
Matches 1049; Conservative 0; Mismatches 0; Indels 91; Gaps 2;

QY 1 MAWKTLPIYLLLLLSVFVIQVSSQ-----25
DB 1 MAWKTLPIYLLLLLSVFVIQVSSQ-----60
QY 26 -----ELSKGRCFBSFERGECDDAOCKKYDKCCPDYESFCA-----64
DB 61 KRVCYTAELSKGRCFBSFERGECDDAOCKKYDKCCPDYESFCAEVHNPTSPPSKKAP 120
QY 65 -----BEHSVSENQESSSSSSSSSSSTI 89
DB 121 PPSGASQTIKSTTKRSPKPNKKTKKVIESEITEHSVSENQESSSSSSSSSTI 180
QY 90 KIKSSKNSAANRELOKLVKONKNRTKKTKTKPPVVDVDEAGSLDNGDFKVTTPDTST 149
DB 181 KIKSSKNSAANRELOKLVKONKNRTKKTKTKPPVVDVDEAGSLDNGDFKVTTPDTST 240
QY 150 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSUTVNKETTIVETKETTITNKQSTDG 209
DB 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSUTVNKETTIVETKETTITNKQSTDG 300
QY 210 KBKTTSAKETQSIKETSADLAPTSKVLAKPTPKAFTTTKGPALITPKETPTTTPKEPAS 269
DB 301 KBKTTSAKETQSIKETSADLAPTSKVLAKPTPKAFTTTKGPALITPKETPTTTPKEPAS 360
QY 270 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTTKEP 329
DB 361 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTKPEAPTTKPEAPTTTKEP 420
QY 330 APTTTKSAPTTKPEAPTTPKKPAFTTTPKEAPTTKPEAPTTKPEAPTTTKEPAPT 389
DB 421 APTTTKSAPTTKPEAPTTPKKPAFTTTPKEAPTTKPEAPTTKPEAPTTTKEPAPT 480
QY 390 EPAPTAPKPAFTTTPKEAPTTKPEAPTTTKEPSPTTTPKEAPTTTKEPAPT 449
DB 481 EPAPTAPKPAFTTTPKEAPTTKPEAPTTTKEPSPTTTPKEAPTTTKEPAPT 540
QY 450 TTKSAPTTKPEPSPTTTPKEAPTTKPEAPTTKPEAPTTKPEAPTTTKEPAPT 509
DB 541 TTKSAPTTKPEPSPTTTPKEAPTTKPEAPTTKPEAPTTKPEAPTTTKEPAPT 600
QY 510 APTAPKEAPTTPKETAPTTPKLTPTTPEKLAFTTTPKEAPTTPEELAPTTPEEPTPT 569
DB 601 APTAPKEAPTTPKETAPTTPKLTPTTPEKLAFTTTPKEAPTTPEELAPTTPEEPTPT 660
QY 570 PEEPAPTTPKAAAPNTPKPEAPTTTPKEAPTTTPKEAPTTTPKEAPTTTPKEAPTTTKEP 629

```

## RESULT 8

```

US-10-124-557-2
; Sequence 2, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luan
; REGISTRATION NUMBER: 31-822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:

```

```

661 PEEPAPTTPKAAAPNTPKPEAPTTTPKEAPTTTPKEAPTTTPKEAPTTTPKEAPTTTPKEP 720
QY 630 APTTPKKPAPKELAPTTTKEPTTTSKAPATTTPKGTAPTTTPKEAPTTTPKEAPTTTPKG 689
DB 721 APTTPKKPAPKELAPTTTKEPTTTSKAPATTTPKGTAPTTTPKEAPTTTPKEAPTTTPKG 780
QY 690 TAPTTLKEAPATTTPKKPAPKELAPTTTKEPTTTSKAPATTTPKGTAPTTTPKEAPTTTPKE 749
DB 781 TAPTTLKEAPATTTPKKPAPKELAPTTTKEPTTTSKAPATTTPKGTAPTTTPKEAPTTTPKE 840
QY 750 KPAPTTPPETPPPTTSVSTPTTTHKSPDESTPELSAPTPKALENSPKEGCVPT 809
DB 841 KPAPTTPPETPPPTTSVSTPTTTHKSPDESTPELSAPTPKALENSPKEGCVPT 900
QY 810 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETAATTTKTTESKITATTQV 869
DB 901 TKTPAATKPEMTTAKDKTTERDLRTTPETTTAAPKMTKETAATTTKTTESKITATTQV 960
QY 870 TSTTTODTTPEKITTLTKTTTLAPKVTTKKTTTTEIMNKPETAKPKDRATNSKATTPK 929
DB 961 TSTTTODTTPEKITTLTKTTTLAPKVTTKKTTTTEIMNKPETAKPKDRATNSKATTPK 1020
QY 930 PQKPTKAPKKPTSTKKPKTMRVRKPKTTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 989
DB 1021 PQKPTKAPKKPTSTKKPKTMRVRKPKTTTPTRKMTSTMPELNPTSRIAEAMLQTTTRPN 1080
QY 990 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVRVFNQGIINPMLS 1049
DB 1081 QTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMDYLPVRVFNQGIINPMLS 1140

```





[illegible]

Db 121 KPTKPPVWDAGSLDNGDFKVTTPDSTTQHNKVSFSPKITTAKPINRPSLPPNSDT 180  
QY 180 SKETSLTVNKETTVETKTTTTNKTSTGKEKTTSAKETQSIKTSKOLAPTSKVLAK 239  
Db 181 SKETSLTVNKETTVETKTTTTNKTSTGKEKTTSAKETQSIKTSKOLAPTSKVLAK 240  
QY 240 PTPKAEITTKGPAITTKPEPTTKPEPASITPKPEPTTTIKSAPTTKPEAPTTTKSAP 299  
Db 241 PTPKAEITTKGPAITTKPEPTTKPEPASITPKPEPTTTIKSAPTTKPEAPTTTKSAP 300  
QY 300 TTPKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPE 359  
Db 301 TTPKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTTTKPE 360  
QY 360 PAPTTPKEPTTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 419  
Db 361 PAPTTPKEPTTTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 420  
QY 420 TKERSPTTKPEAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAP 479  
Db 421 TKERSPTTKPEAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAP 480  
QY 480 TTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 539  
Db 481 TTPKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTT 540  
QY 540 KLAAPTPEKPAITTEELAPTTPEBPTTTPPEAPTTTPKAAAPNTPEAPTT 599  
Db 541 KLAAPTPEKPAITTEELAPTTPEBPTTTPPEAPTTTPKAAAPNTPEAPTT 600  
QY 600 TTPKPEAPTTKETAPTTPKGTAPTTLKPEAPTTPKKPAPKELAPTTKETSDKPA 659  
Db 601 TTPKPEAPTTKETAPTTPKGTAPTTLKPEAPTTPKKPAPKELAPTTKETSDKPA 660  
QY 660 PTPPKGTAPTTKPEAPTTPKGTAPTTLKPEAPTTPKKPAPKELAPTTKGP 719  
Db 661 PTPPKGTAPTTKPEAPTTPKGTAPTTLKPEAPTTPKKPAPKELAPTTKGP 720  
QY 720 TSTTSKPAPTTKETAPTTPKPEAPTTPKKPAPTTPEPTTPPTSEVSTPTTKPTTIH 779  
Db 721 TSTTSKPAPTTKETAPTTPKPEAPTTPKKPAPTTPEPTTPPTSEVSTPTTKPTTIH 780  
QY 780 KSPDSTPELSAPETPKALENSPKPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPT 839  
Db 781 KSPDSTPELSAPETPKALENSPKPGVPTTKTPAATKPEMTTTAKDKTTERDLRTTPT 840  
QY 840 TTAAPKMTKETATTTETKTTESKITATTTQVSTTTQDITTPFKITLTKTTLAPKVTITKK 899  
Db 841 TTAAPKMTKETATTTETKTTESKITATTTQVSTTTQDITTPFKITLTKTTLAPKVTITKK 900  
QY 900 TTTTTEIMNKPEATAKPKORATNSKATTPKPKPKAPKPKPTSTKPKMTPVRPKPTTP 959  
Db 901 TTTTTEIMNKPEATAKPKORATNSKATTPKPKPKAPKPKPTSTKPKMTPVRPKPTTP 960  
QY 960 TPKMTSTPELNPSTSRIAEMLQTTTRNQTPNSKLVNPKSEDAGGAEGETPHMLR 1019  
Db 961 TPKMTSTPELNPSTSRIAEMLQTTTRNQTPNSKLVNPKSEDAGGAEGETPHMLR 1020  
QY 1020 PHVFMPEVTPDMDYLRPVNQGIINPMLS 1049  
Db 1021 PHVFMPEVTPDMDYLRPVNQGIINPMLS 1050

RESULT 11  
US-10-124-557-46  
; Sequence 46, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.

; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:  
US-10-124-557-46

Query Match 96.7%; Score 5393.3; DB 13; Length 1320;  
Best Local Similarity 96.1%; Pred. No. 9.3e-135;  
Matches 1015; Conservative 8; Mismatches 26; Indels 7; Gaps 1;  
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCKGRCSFESFERGRCDCDQAQCKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQELSCKGRCSFESFERGRCDCDQAQCKYDKCCPDYE 60  
QY 61 SFCAEBHSVSNQESSSSSSSSSTTWIKSSKNNSAANRELQKL-----KVQDNK 113  
Db 61 SFCAEVHNPSTSPFSSKAPPPSGASQTKSTTKRSPKPPNKKTKKVFSEBITEYKDNK 120  
QY 114 KNRTKKPTPKPPVVDAGSLDNGDFKVTTPDSTTQHNKVSFSPKITTAKPINRPSL 173  
Db 121 KNRTKKPTPKPPVVDAGSLDNGDFKVTTPDSTTQHNKVSFSPKITTAKPINRPSL 180  
QY 174 PPNSTSKETSLTVNKETTVETKTTTTNKTSTGKEKTTSAKETQSIKTSKOLAPT 233  
Db 181 PPNSTSKETSLTVNKETTVETKTTTTNKTSTGKEKTTSAKETQSIKTSKOLAPT 240  
QY 234 SKVLAKPTPKAEITTKGPAITTKPEPTTKPEPASITPKPEPTTTIKSAPTTKPEAP 293  
Db 241 SKVLAKPTPKAEITTKGPAITTKPEPTTKPEPASITPKPEPTTTIKSAPTTKPEAP 300  
QY 294 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTT 353  
Db 301 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTT 360  
QY 354 PTPKPEAPTTPKPEPTTKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 413



Db 601 PKPAPPTPKGPAPTTPKETAATTPKGNAPITLKEPAPTTPKPKAPKELAPPTTKPTST 660  
QY 654 TSDKAPPTPKGTAPTTPKEPAPTTPKBPAPTTPKGTAPTTLKEPAPTTPKPKAPKELAP 713  
Db 661 TSDKAPPTPKGTAPTTPKEPAPTTPKBPAPTTPKGTAPTTLKEPAPTTPKPKAPKELAP 720  
QY 714 TTTKGTSTTSKAPAPTTPKETAPTTPKEPAPTTPKBPAPTTPKGTAPTTLKEPAPTTPKPKAPKELAP 773  
Db 721 TTTKGTSTTSKAPAPTTPKETAPTTPKEPAPTTPKBPAPTTPKGTAPTTLKEPAPTTPKPKAPKELAP 780  
QY 774 EPTTHKSPDESTPELSAEPPTKALENSPKEGVPPTTKPAATKPEMTTAKDKTERDL 833  
Db 781 EPTTHKSPDESTPELSAEPPTKALENSPKEGVPPTTKPAATKPEMTTAKDKTERDL 840  
QY 834 RTTPETTTAAAPKMTKETATTTKTESKITATTTQVTSSTTTQDTPPKITTLKTTLAPK 893  
Db 841 RTTPETTTAAAPKMTKETATTTKTESKITATTTQVTSSTTTQDTPPKITTLKTTLAPK 900  
QY 894 VTTTKKTTITTTIMNKPEATKPKDRATNSKATTPKPKPTKAPKPTSTKPKTMPVR 953  
Db 901 VTTTKKTTITTTIMNKPEATKPKDRATNSKATTPKPKPTKAPKPTSTKPKTMPVR 960  
QY 954 KPCTTTPKMTSTMPELNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPXSEDAAGAGET 1013  
Db 961 KPCTTTPKMTSTMPELNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPXSEDAAGAGET 1020  
QY 1014 PHEMLRPHVMPVETPDMDLPRVPNOGIINPMLS 1049  
Db 1021 PHEMLRPHVMPVETPDMDLPRVPNOGIINPMLS 1056

RESULT 13  
US-10-124-557-40  
; Sequence 40, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JUN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190

TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1361 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:  
US-10-124-557-40  
  
Query Match 96.5%; Score 5379.2; DB 13; Length 1361;  
Best Local Similarity 92.5%; Pred. No. 2.3e-134;  
Matches 1015; Conservative 8; Mismatches 26; Indels 48; Gaps 2;  
  
QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25  
Db 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 60  
  
QY 26 -----ELSCKGRCPESFERGECDCDAQCKKYDKCCPDYESFCABEHSVSENQSSSS 79  
Db 61 KKVCTAELSCKGRCPESFERGECDCDAQCKKYDKCCPDYESFCABEHSVSENQSSSS 120  
  
QY 80 SSSSSSTIWKIKSSKNKSAANRELQKL-----KVKNKKKRTKKKTPKPPVVDK 132  
Db 121 PPSGASQTIKSTTKRSPKPPKKTKKVIKSEBETEVKDKKKRTKKKTPKPPVVDK 180  
  
QY 133 SGLDNGDFKVTTPDSTTQHNVKVSIPKLTAKPINRPSLPNSDTSKETSLSLVNKEIT 192  
Db 181 SGLDNGDFKVTTPDSTTQHNVKVSIPKLTAKPINRPSLPNSDTSKETSLSLVNKEIT 240  
  
QY 193 VETKETTTNNKOTS TDGKEKTTSAKETQSIETSAKDAPTSKVLAKPTPKAETTTKGA 252  
Db 241 VETKETTTNNKOTS TDGKEKTTSAKETQSIETSAKDAPTSKVLAKPTPKAETTTKGA 300  
  
QY 253 LTTKPEPTTPKPEASTTPKPEPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 312  
Db 301 LTTKPEPTTPKPEASTTPKPEPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 360  
  
QY 313 PAPTTPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTPKKAPPTTPKPEAPTTTKEPT 372  
Db 361 PAPTTPKPEAPTTTKPEAPTTTKSAPTTPKPEAPTTPKKAPPTTPKPEAPTTTKEPT 420  
  
QY 373 PKPEAPTTKEPAPTTPKPEAPTTAPKAPPTTPKPEAPTTTPKPEAPTTTKEPSPPTTKEPA 432  
Db 421 PKPEAPTTKEPAPTTPKPEAPTTAPKAPPTTPKPEAPTTTPKPEAPTTTKEPSPPTTKEPA 480  
  
QY 433 PTTTKGAPTTKPEAPTTTKSAPTTPKPEAPTTTPKPEAPTTTPKPEAPTTTKEPAPTTTKE 492  
Db 481 PTTTKGAPTTKPEAPTTTKSAPTTPKPEAPTTTPKPEAPTTTPKPEAPTTTKEPAPTTTKE 540  
  
QY 493 PAPTTPKPEAPTTTKKAPPTAPKPEAPTTTPKETAATTPKELTPTTPEKLAAPTTPKEPAPT 552  
Db 541 PAPTTPKPEAPTTTKKAPPTAPKPEAPTTTPKETAATTPKELTPTTPEKLAAPTTPKEPAPT 600  
  
QY 553 TPBELAPTTPEEPTTPTPPEAPPTPKAAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTKE 612  
Db 601 TPBELAPTTPEEPTTPTPPEAPPTPKAAAPNTPKPEAPTTTPKPEAPTTTPKPEAPTTTKE 660  
  
QY 613 TAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKAPPTTPKGTAPTTPK 672  
Db 661 TAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKAPPTTPKGTAPTTPK 720  
  
QY 673 EPAPTTPKPEAPTTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKAPPTTP 732  
Db 721 EPAPTTPKPEAPTTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTTKEPTSTTSKAPPTTP 780  
  
QY 733 KETAPTTPKPEAPTTTPKPKAPPTTPETPTTSEVSTPTTTPKPTTIHKSDDESTPELSAE 792  
Db 781 KETAPTTPKPEAPTTTPKPKAPPTTPETPTTSEVSTPTTTPKPTTIHKSDDESTPELSAE 840  
  
QY 793 PTPKALENSPKPEGVPPTTKPAATKPEMTTAKDKTERDLRTTPTTETTTAAAPKMTKETAT 852

Db 841 PTPKALENSKPEGVPTTKTPAATKPEMTTAKDKTTERDLRTTPEITTAAPKMTKETAT 900  
QY 853 TTEKTTESKITATTQVTSITTQDTPPKLITLTKTTTLAPKVTTKKTIITTIMNKPEE 912  
Db 901 TTEKTTESKITATTQVTSITTQDTPPKLITLTKTTTLAPKVTTKKTIITTIMNKPEE 960  
QY 913 TAKPKDRATNSKATTPKPKQPTKAPKKPTSTKKPKTMPRVRKPKTTPTRKMTSTMPELN 972  
Db 961 TAKPKDRATNSKATTPKPKQPTKAPKKPTSTKKPKTMPRVRKPKTTPTRKMTSTMPELN 1020  
QY 973 PTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMD 1032  
Db 1021 PTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSEDAGGAGETPHMLLRPHVFMPEVTPDMD 1080  
QY 1033 YLPRVENQGIINPMLS 1049  
Db 1081 YLPRVENQGIINPMLS 1097

RESULT 14  
US-10-124-557-74  
; Sequence 74, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 74:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1038 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:

Query Match 96.1%; Score 5358.7; DB 13; Length 1038;  
Best Local Similarity 95.9%; Pred. No. 5.8e-134;  
Matches 1006; Conservative 0; Mismatches 0; Indels 43; Gaps 1;  
QY 1 MAWKTLPIYLLLLLSVFVIOQVSSQELSCKGRCFESFERGREGDCDCAOCKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLLSVFVIOQVSSQELSCKGRCFESFERGREGDCDCAOCKYDKCCPDYE 60  
QY 61 SFCAEHSVSENQESSSSSSSSSSSIWKIKSSKNSAANRELQKKLVKDNKNRITKK 120  
Db 61 SFCAE-----VKDNKNRITKK 77  
QY 121 PTPKPPVDEAGSLGNDPKVTPDTSITTOHNVKSTSPKITTAKINPRPSLPNNSDTS 180  
Db 78 PTPKPPVDEAGSLGNDPKVTPDTSITTOHNVKSTSPKITTAKINPRPSLPNNSDTS 137  
QY 181 KETSLTVNKETTIVETKETTITNKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKP 240  
Db 138 KETSLTVNKETTIVETKETTITNKQTSIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKP 197  
QY 241 TPKAETTTKGPALTTTPKEPTTPPKEPASTTPKEPTPTTIKSAPTTPKEPATTTKSAPT 300  
Db 198 TPKAETTTKGPALTTTPKEPTTPPKEPASTTPKEPTPTTIKSAPTTPKEPATTTKSAPT 257  
QY 301 TPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEP 360  
Db 258 TPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEP 317  
QY 361 APPTTPKEPTTPPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTT 420  
Db 318 APPTTPKEPTTPPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTT 377  
QY 421 KEPSPTTPKEPATTTTKSAPTTTTKSAPTTTTKSAPTTTTKSAPTTTTKSAPTTTTKSA 480  
Db 378 KEPSPTTPKEPATTTTKSAPTTTTKSAPTTTTKSAPTTTTKSAPTTTTKSAPTTTTKSA 437  
QY 481 TPKKAPATTTTPKEPATTTTPKEPATTTTKKAPATTPKEPATTTTPKEPATTTTPKEPAT 540  
Db 438 TPKKAPATTTTPKEPATTTTPKEPATTTTKKAPATTPKEPATTTTPKEPATTTTPKEPAT 497  
QY 541 LAPTTPEKAPATTPBELAPTTPEEPPTTPPEPATTTPKAAAPNTPKEPATTTPKEPAT 600  
Db 498 LAPTTPEKAPATTPBELAPTTPEEPPTTPPEPATTTPKAAAPNTPKEPATTTPKEPAT 557  
QY 601 TPKEPATTTPKETAPTTPKGATPTTLKEPATTTPKKAPKELAPITTKKEPTSTTSKAP 660  
Db 558 TPKEPATTTPKETAPTTPKGATPTTLKEPATTTPKKAPKELAPITTKKEPTSTTSKAP 617  
QY 661 TTPKGATPTTPKEPATTTTPKEPATTTTPKGATPTTLKEPATTTPKKAPKELAPITTKGPT 720  
Db 618 TTPKGATPTTPKEPATTTTPKEPATTTTPKGATPTTLKEPATTTPKKAPKELAPITTKGPT 677  
QY 721 STTSKAPATTPPKETAPTTPKEPATTTPKKAPATTTPEPPTTSEVSTPTTKKEPTTIHK 780  
Db 678 STTSKAPATTPPKETAPTTPKEPATTTPKKAPATTTPEPPTTSEVSTPTTKKEPTTIHK 737  
QY 781 SPDESTPELSAEPKALENGKPEGVPTTKPAATKPEMTTTAKDKTTERDLRTTPEIT 840  
Db 738 SPDESTPELSAEPKALENGKPEGVPTTKPAATKPEMTTTAKDKTTERDLRTTPEIT 797  
QY 841 TAAPKMTKETATTTTEKTTESKITATTQVTSITTQDTPPKLITLTKTTTLAPKVTTKKTI 900  
Db 798 TAAPKMTKETATTTTEKTTESKITATTQVTSITTQDTPPKLITLTKTTTLAPKVTTKKTI 857  
QY 901 ITTTEIMNKPEETAKPKDRATNSKATTPKPKQPTKAPKKPTSTKKPKTMPRVRKPKTTP 960  
Db 858 ITTTEIMNKPEETAKPKDRATNSKATTPKPKQPTKAPKKPTSTKKPKTMPRVRKPKTTP 917  
QY 961 PRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSEDAGGAGETPHMLLRP 1020  
Db 918 PRKMTSTMPELNPTSRIAEAMLQTTTRPNQTPNSKLVENVNPKSEDAGGAGETPHMLLRP 977  
QY 1021 HVFMPEVTPDMDYLPRVENQGIINPMLS 1049



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 115.337 Seconds  
(without alignments)  
5233.063 Million cell updates/sec

Title: SEQ1-D  
Perfect score: 5576  
Sequence: 1 MAWKTPILVLLLSVFIQ.....DMDYLPRVNPQGIINPMLS 1049

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_02: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5546.9	99.5	1404	2	Q92954
2	5523.9	99.1	1404	2	Q9BX49
3	3143	56.4	933	2	Q62M25
4	3143	56.4	933	2	BAD18580
5	2445.4	43.9	1054	2	Q9JW99
6	1346.8	24.2	5179	1	MUC2_HUMAN
7	1320.6	23.7	1225	2	Q9VR49
8	1314.2	23.6	1761	2	Q7KTF6
9	1243.6	22.3	3150	2	Q7PMD5
10	1243.6	22.3	3150	2	AAS64673
11	1169.4	21.0	3409	2	Q6SSB6
12	1169.4	21.0	3409	2	AAS07044
13	1128.7	20.2	1664	1	SLPI_CLOUTM
14	1113.7	20.0	1349	2	Q8WWQ4
15	1090.2	19.6	3432	2	Q8IR51
16	1090.2	19.6	3458	2	Q8IR52
17	1080	19.4	3889	2	Q6SSB8
18	1080	19.4	3889	2	AAS07042
19	1068	19.2	1795	2	Q76894
20	1051.3	18.9	1079	2	Q9N4S7
21	1034.6	18.6	9234	2	Q7KTP5
22	1034.6	18.6	9234	2	AAN10531
23	1020.5	18.3	2284	2	Q9VPG1
24	1020.4	18.3	972	2	Q7QKK7
25	1012.6	18.2	1607	2	Q8H6Q5
26	1012.6	18.2	1607	2	AAP74661
27	1002.8	18.0	1274	2	Q20007
28	993.7	17.8	1489	2	Q96449
29	990.9	17.8	10578	2	Q8ISF5
30	989.5	17.7	23015	2	Q8IQ48
31	989.5	17.7	23015	2	AAN10358

## RESULT 1

Q92954 PRELIMINARY; PRT; 1404 AA.  
AC Q92954;  
DT 01-FEB-1997 (TrEMBLrel. 02, Created)  
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Megakaryocyte stimulating factor.  
OS Megakaryocyte (Human).  
OC Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
RT "Purification, Biochemical Characterization, and Cloning of a Novel  
RT Megakaryocyte Stimulating Factor that has Megakaryocyte Colony  
RT Stimulating Activity.";  
RL Blood 78:279-279(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,  
RA Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,  
RA Jacobs K., Turner K.;  
RT "A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";  
RL (In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,  
RL Mosher D.F. (eds.);  
RL BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier  
RL Science Publishers B.V. (1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,  
RA Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,  
RA Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,  
RA Bhatia S., Kriz R., Hewick R., Clark S.C.;  
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U701336; AAB09089.1; -;  
DR HSSP; P04004; 10C0.  
DR Genew; HGNC:9364; PRG4.  
DR GO; GO:0008283; P:cell proliferation; TAS.  
DR InterPro; IPR000585; Hemopexin.  
DR InterPro; IPR001212; Somatomedin\_B.  
DR Pfam; PF00045; Hemopexin; 2.  
DR PRINTS; PR00022; Somatomedin\_B; 2.  
DR SMART; SM00120; HX; 2.  
DR SMART; SM00120; SO; 2.  
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.  
DR PROSITE; PS00524; SOMATOMEDIN B; 2.  
SQ SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;

32 988.9 17.7 18519 2 Q8ISF6  
33 988.9 17.7 18534 2 Q8ISF7  
34 988.4 17.7 34350 2 Q8WZ42  
35 988 17.7 7962 2 Q10465  
36 978.7 17.6 5703 1 MUSB\_HUMAN  
37 975.9 17.5 2112 2 Q9VEL9  
38 974.8 17.5 2187 2 P70670  
39 964.4 17.3 926 2 Q9VVG2  
40 960.4 17.2 1720 2 Q8I486  
41 956.8 17.2 1480 2 Q9LIE8  
42 956.7 17.2 1458 2 Q757N5  
43 956.7 17.2 1458 2 Q757N5  
44 953.5 17.1 1121 2 AAS52662  
45 945.1 16.9 2042 2 Q767L8

## ALIGNMENTS





```

QY 390 EPAPTAPKCAPPTPKBPAPPTPKBPAPPTTKPSPTTKBPAPPTTKSAPTTTKBPAPT 449
Db 481 EPAPTAPKCAPPTPKBPAPPTPKBPAPPTTKPSPTTKBPAPPTTKSAPTTTKBPAPT 540
QY 450 TTKSAPPTPKPSPTTKBPAPPTPKBPAPPTPKBPAPPTTKBPAPPTTKBPAPT 509
Db 541 TTKSAPPTPKPSPTTKBPAPPTPKBPAPPTPKBPAPPTTKBPAPPTTKBPAPT 600
QY 510 APTAPKBPAPPTPKBPAPPTPKBPAPPTTKBPAPPTTKBPAPPTTKBPAPT 569
Db 601 APTAPKBPAPPTPKBPAPPTPKBPAPPTTKBPAPPTTKBPAPPTTKBPAPT 660
QY 570 PEPAPPTPKAAPNTPKBPAPPTPKBPAPPTTKBPAPPTTKBPAPPTTKBPAPT 629
Db 661 PEPAPPTPKAAPNTPKBPAPPTPKBPAPPTTKBPAPPTTKBPAPPTTKBPAPT 720
QY 630 APTTEKBPAPKELAPPTTKPSPTTKBPAPPTPKBPAPPTTKBPAPPTTKBPAPT 689
Db 721 APTTPKBPAPKELAPPTTKPSPTTKBPAPPTPKBPAPPTTKBPAPPTTKBPAPT 780
QY 690 TAPTTLLKEPAPPTPKBPAPKELAPPTTKPSPTTKBPAPPTTKBPAPPTTKBPAPT 749
Db 781 TAPTTLLKEPAPPTPKBPAPKELAPPTTKPSPTTKBPAPPTTKBPAPPTTKBPAPT 840
QY 750 KPAPTTPPTPPTTSVSTPTTKBPAPPTTKBPAPPTTKBPAPPTTKBPAPT 809
Db 841 KPAPTTPPTPPTTSVSTPTTKBPAPPTTKBPAPPTTKBPAPPTTKBPAPT 900
QY 810 TKTPAAKTPMTTAKDKTERDLRTPTTAAAPKMTKATATTKTTESKITATTQV 869
Db 901 TKTPAAKTPMTTAKDKTERDLRTPTTAAAPKMTKATATTKTTESKITATTQV 960
QY 870 TSTTTQDTTTPFKITLTKTTLLAPKVTTKKTIITTEIMNKPESTAKPKDRATNSKATTPK 929
Db 961 TSTTTQDTTTPFKITLTKTTLLAPKVTTKKTIITTEIMNKPESTAKPKDRATNSKATTPK 1020
QY 930 POKPTKAPKPTSTKPKTTPRVKPKTTTTPRKMSTSTMPNLNPTSRIAEAMLQTTTRN 989
Db 1021 POKPTKAPKPTSTKPKTTPRVKPKTTTTPRKMSTSTMPNLNPTSRIAEAMLQTTTRN 1080
QY 990 QTPNSKLVNPKSEDAGAGETHMMLRPHVFWPEVTPDMDYLPVFNQGIINPMLS 1049
Db 1081 QTPNSKLVNPKSEDAGAGETHMMLRPHVFWPEVTPDMDYLPVFNQGIINPMLS 1140

RESULT 3
Q6ZMZ5 PRELIMINARY; PRT; 933 AA.
AC Q6ZMZ5;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein FLJ16561.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Syncovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuno Y., Nagai K., Isogai T.;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131434; BAD18580.1; -
DR InterPro; IPR000585; Hemoexin.
DR InterPro; IPR001212; Somatomedin_B.
DR Pfam; PF00045; Hemoexin; 2.

```

```

DR Pfam; PF01033; Somatomedin_B; 1.
DR PRINTS; PR00022; SOMATOMEDINB.
DR SMART; SMO0120; HX; 2.
DR SMART; SMO0201; SO; 1.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.
DR PROSITE; PS00524; SOMATOMEDIN_B; 1_1.
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match
Best Local Similarity 56.4%; Score 3143; DB 2; Length 933;
Matches 618; Conservative 0; Mismatches 1; Indels 480; Gaps 2;

QY 1 MANKTLPYILLILLVFIQOVSSQELSCRGCFESPERGREGCDCAQCKKYKCCPDYE 60
Db 1 MANKTLPYILLILLVFIQOVSSQELSCRGCFESPERGREGCDCAQCKKYKCCPDYE 60
QY 61 SFCA-----EHSVS 70
Db 61 SFCAEVHNPTSPSSSKAPPPSGASQTIKSTTKRSPKPNKKTKKVIESEITEHSVS 120
QY 71 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKKLVKDNKKNRTKKKPTPKPPVDE 130
Db 121 ENQESSSSSSSSSTIRKIKSSKNSAANRELQKKLVKDNKKNRTKKKPTPKPPVDE 180
QY 131 AGSGLDNGDPKVTTPDTSTTOHNKUSTSPKITTAKINPRPSLPNSDTSKETSITVKE 190
Db 181 AGSGLDNGDPKVTTPDTSTTOHNKUSTSPKITTAKINPRPSLPNSDTSKETSITVKE 240
QY 191 TVVETKETTITNKQTSIDGKEKTSKETSQTSKTSKADLAPTSKVLAKPTPKAETTKG 250
Db 241 TVVETKETTITNKQTSIDGKEKTSKETSQTSKTSKADLAPTSKVLAKPTPKAETTKG 300
QY 251 PALTTPKEPTPTPKBPASTTPKEPTPTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTT 310
Db 301 PALTTPKEPTPTPKBPASTTPKEPTPTTIKSAPTTPKBPAPTTTKSAPTTPKBPAPTTT 360
QY 311 KEPAPTTPKEPAPTTTKBPAPTTKSAPTTPKBPAPTTTPKBPAPTTPKBPAPTT 370
Db 361 KEPAPTPKB----- 370
QY 371 TTPKEPAPTTKEPAPTTTPKBPAPTAPKBPAPTTTPKEPAPTTTPKEPAPTTTPKE 430
Db 371 ----- 370
QY 431 PAPTTPKSAPTTKEPAPTTTKSAPTTTPKEPSFTTKBPAPTTTPKBPAPTTTPKBPAPT 490
Db 371 ----- 370
QY 491 KEPAPTTPKEPAPTTTKBPAPTAPKBPAPTTTPKETAPTTPKLTPTTPEKLAPTTPEKA 550
Db 371 ----- 370
QY 551 FTTPEELAPTTPEPTPTTPPEPAPTTTPKAAAPNTPKBPAPTTTPKBPAPTTTPKEPAPTT 610
Db 371 ----- 370
QY 611 KETAPTTPKGTAPTTLKEPAPTTTPKBPAPKELAPTTTKETSTTTSDKAPTTTPKGTAPT 670
Db 371 ----- 370
QY 671 PKEPAPTTPKBPAPTTTPKGTAPTTLKEPAPTTTPKBPAPKELAPTTTKGPTSTTSKAPAPT 730
Db 371 ----- 370
QY 731 TPKETAPTTPKBPAPTTTPKBPAPTTPEPTTSEVSTPTTKETPTTIHKSPDSTPELS 790
Db 371 -----PAPTTPEPTTSEVSTPTTKETPTTIHKSPDSTPELS 410
QY 791 APTPKALENSPKPGVPTTKTPAATKPEMTTTAKDKTERDLRTTTPETTTAAAPKMTKET 850
Db 411 APTPKALENSPKPGVPTTKTPAATKPEMTTTAKDKTERDLRTTTPETTTAAAPKMTKET 470
QY 851 ATTTEKTTESKITATTQVTTSTTQDTTTPPKITLTKTTLLAPKVTTKKTIITTEIMNK 910

```

```

Db 471 ATTEKTTESKITATTTQVTSITTTQDTTPPKIITLKTTLAPKVTTTKKTIITTEIMNKP 530
QY 911 EETAKPKDRATNSKATTPKPKQKTPKAPKPKTSTKKPKTMPRVRKPKTTTPRKMTSTWPE 970
Db 531 EETAKPKDRATNSKATTPKPKQKTPKAPKPKTSTKKPKTMPRVRKPKTTTPRKMTSTWPE 590
QY 971 LNPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 1030
Db 591 LNPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 650
QY 1031 MDYLPRVNOGIIINPMLS 1049
Db 651 MDYLPRVNOGIIINPMLS 669

RESULT 4
BADI8580 PRELIMINARY; PRT; 933 AA.
AC BADI8580;
DT 12-MAY-2004 (TrEMBLrel. 27, Created)
DT 12-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DT 12-MAY-2004 (TrEMBLrel. 27, Last annotation update)
DE CDNA FLJ16561 fis, Clone SYN04003981, moderately similar to Homo
DE sapiens proteoglycan 4, (megakaryocyte stimulating factor, articular
DE superficial zone protein, campodactyly, arthropathy, coxa vara,
DE pericarditis syndrome) (PRG4).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP TISSUE=Synovial membrane tissue;
RC Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131434; BADI8580.1; -.
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match 56.4%; Score 3143; DB 2; Length 933;
Best Local Similarity 56.2%; Pred. No. 3,7e-54;
Matches 618; Conservative 0; Mismatches 1; Indels 480; Gaps 2;

QY 1 MAWKTLPIYLLLSVFIQVSSQELSCKGRCPESPFGRECDCAQCKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLSVFIQVSSQELSCKGRCPESPFGRECDCAQCKYDKCCPDYE 60
QY 61 SFCA-----EHSVS 70
Db 61 SFCAVHNTPSPSSKAPPPSGASQTIKSTTKSPXPNKKTKKVIIEEITEHSVS 120
QY 71 ENOESSSSSSSSSSSTWIKSKSGAANRELQKLKVDKNKNRTKKXTPKPPVUDE 130
Db 121 ENOESSSSSSSSSSSTWIRKIKSSKNGAANRELQKLKVDKNKNRTKKXTPKPPVUDE 180
QY 131 AGSLDNGDFKVTTPDSTTOHNKVSTSPKLTAKPINRPSLPNSDTSKETSITWKE 190
Db 181 AGSLDNGDFKVTTPDSTTOHNKVSTSPKLTAKPINRPSLPNSDTSKETSITWKE 240
QY 191 TTIVETKTTTNTKQTSIDGKEKTTSAKETQSIETKSADLAPTSGVLAKPTPKAETITKG 250
Db 241 TTIVETKTTTNTKQTSIDGKEKTTSAKETQSIETKSADLAPTSGVLAKPTPKAETITKG 300
QY 251 PALITPKPEPTTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTT 310
Db 301 PALITPKPEPTTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTT 360

```

```

QY 311 KEPAPTTTPKEPAPTTTKEPAPTTTKSAPTTPKEPAPTTTPKBPAPTTTPKEPAPTTTPKEPPTP 370
Db 361 KEPAPTTTPKE----- 370
QY 371 TTPKEPAPTTTTPKEPAPTTTPKEPAPTTAPKBPAPTTTPKEPAPTTTPKEPAPTTTPKEPPTTTPKE 430
Db 371----- 370
QY 431 PAPTTPKGAPTTTPKEPAPTTTKSAPTTPKEPPTTTPKEPAPTTTPKEPAPTTTPKBPAPTTTP 490
Db 371----- 370
QY 491 KEPAPTTTPKEPAPTTTKKBPAPTTAPKBPAPTTPKETAPTTPKKLTTPTPKEKLAPTTTPKEPA 550
Db 371----- 370
QY 551 PTTPELAPTTTPPEPTTPPEPAPTTPKAAANPTPKBPAPTTTPKEPAPTTTPKEPAPTTTP 610
Db 371----- 370
QY 611 KETAPTPKGTAPTTTPKEPAPTTTPKBPAPKELAPTTTKBPTSTTSKBPAPTTTPKGTAPTT 670
Db 371----- 370
QY 671 PKBPAPTTTPKEPAPTTPKGTAPTTTPKEPAPTTTPKBPAPKELAPTTTKGPTSTTSKBPAPT 730
Db 371----- 370
QY 731 TPKETAPTTTPKEPAPTTTPKBPAPTTTPETTPPTTSEVSTPTTTPKEPTTIHKSPDESTPELS 790
Db 371-----PAPTTPETTPPTTSEVSTPTTTPKEPTTIHKSPDESTPELS 410
QY 791 ABPTPKALENSPKBPAPTTTPKPAATKPEMTTAKDKTTERDLRTTPETTTAAPKVKTKET 850
Db 411 ABPTPKALENSPKBPAPTTTPKPAATKPEMTTAKDKTTERDLRTTPETTTAAPKVKTKET 470
QY 851 ATTEKTTESKITATTTQVTSITTTQDTTPPKIITLKTTLAPKVTTTKKTIITTEIMNKP 910
Db 471 ATTEKTTESKITATTTQVTSITTTQDTTPPKIITLKTTLAPKVTTTKKTIITTEIMNKP 530
QY 911 EETAKPKDRATNSKATTPKPKQKTPKAPKPKTSTKKPKTMPRVRKPKTTTPRKMTSTWPE 970
Db 531 EETAKPKDRATNSKATTPKPKQKTPKAPKPKTSTKKPKTMPRVRKPKTTTPRKMTSTWPE 590
QY 971 LNPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 1030
Db 591 LNPTSRIAEAMLOTTTRENQTPNSKLVENVNPKSEDAGAGETPHMLLRPHVFMPEVTPD 650
QY 1031 MDYLPRVNOGIIINPMLS 1049
Db 651 MDYLPRVNOGIIINPMLS 669

RESULT 5
Q9JW99 PRELIMINARY; PRT; 1054 AA.
AC Q9JW99;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus.
GN Name=Pr94;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20573856; PubMed=11124536;
RA Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT "Isolation, characterization and mapping of the mouse and human PRG4
RT (proteoglycan 4) genes.";

```

```

RL Cytogenet. Cell Genet. 90:291-297(2000).
DR EMBL; AB034730; BAA92310.1; -.
DR HSSP; P04004; LOCO.
DR MGI; MGI:1891344; Pr94.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR01212; Somatomedin B.
DR Pfam; PF00045; Hemopexin; 2.
DR Pfam; PF01033; Somatomedin B; 2.
DR PRINTS; PR00022; SOMATOMEDIN.
DR SMART; SM00120; HX; 2.
DR SMART; SM00201; SO; 2.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.
DR PROSITE; PS00524; SOMATOMEDIN B; 2.
DR SEQUENCE 1054 AA; 115991 MW; 4FC64BFA42283235 CRC64;

Query Match 43.9%; Score 2445.4; DB 2; Length 1054;
Best Local Similarity 43.8%; Pred. No. 1.8e-40;
Matches 505; Conservative 37; Mismatches 139; Indels 476; Gaps 22;

QY 1 MAWKTLPIYLLLSLVFVIQVSSQ----- 25
Db 1 MGWKILPVCLSLLLPVVLIQVSSQDLSSCAGRCGEGYSDATCNDYNCQHYMECCDPF 60
QY 26 -----ELSCGRCFSEFGRCCDDAQQKKYDKCPDYSEFC----- 63
Db 61 KRVCGPELSCGRCFSEFGRCCDDSOCKQYKGCADYDFCEEVHNSSTPSSKTAPT 120
QY 64 -----AEHSVSENQESSSSSSSSSSSSSSSSSIWIKI 91
Db 121 PAGASDTIKSTYRSPKSPSTTRTIKVSEEEUTEHSDSENQEP--SSSSSSSSSSSTIRKI 178
QY 92 KSSKNSAANRELQKKLVKDNKNRTKKPTPKPPVVDAGSLDNGDFKVT--TPDTST 149
Db 179 KSSKNS-ANRELQNPVNDKNKNTKKXNPPEPPAVDAGSLDNGEFLTPPPDPPT 237
QY 150 TQHNKVSFPIITAKPINRPSLPNSDTSKETSITVNETTVEPKETITTTNKQSTDG 209
Db 238 TPHSKVATSPKTTAAKPVTPKPSLAPNSDTSKEASLASNKETTVEPKETITTTNKQSSA-S 296
QY 210 KEKTSKETSQIETKSADLAPTSKVLAKPTPKAETTTKGALITPKPEPTTTPKEPAS 269
Db 297 KKITSVETRAEKISDKVDEPT----- 321
QY 270 TTPKEPTTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTT-----PKEP 321
Db 322 -----TTPKNSAPTTTKKPVTTTKESKFLDLPQEP 351
QY 322 APTTTKEAPTTKSAPTTPKEPAPTTPKKAPTTPKGAPTTPKPEPTTTTPKEAPATTK 381
Db 352 EPTTAKEPPTTKKPEPTTRKEPEPTTPKEPEPTTPKEPEPTTPKEPEPTTPKEPTTK 411
QY 382 EPAPTTKEAPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 441
Db 412 KPPEPTTKEPPTTKEPEPTTTKEPEPTTTKEPEPTTTKEPEPTTTKEPEPTTTKEPE 461
QY 442 TTKGAPTTTKSAPTTKEPEPTTTKEPEPTTTKEPEPTTTKEPEPTTTKEPEPTTTKEPE 501
Db 462 -----EPTTKEPEPTTLKEPEPTTPKEPEPTTTKEPEPTTTKEPEPTTTKEPEPT 509
QY 502 APTTTKKAPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 561
Db 510 EPT----- 512
QY 562 PEEPTPTTPEBPAPTTPKAAPNTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAP 621
Db 513 -----TPKEPEPTTPKEPEPTTTKEPEPTTTKEPEPTTTKEPEPTTTKEPEPT 538
QY 622 APTTLKEAPPTPKKAPAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 681
Db 539 -----KEPEPTTPKKP-----EPTTKEPEPTTTKEPEPTTTKEPEPTTTKEPE 564
QY 682 PAPTTPKGTAPTTLKEAPAPTTKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAPKAP 741

```

---

```

Db 565 PEPTTP-----KEPEPTTPKEPE-----TTTKKEPEPTTPKEPEPTTPK 603
QY 742 EPAPTTPKKAPAPTTTPPTTSEVSTPTTTKBTFTTHKSPDBSTPSELSAEPPTKALENS 801
Db 604 EPEPTTPKKPEPTT----- 617
QY 802 PKEPGVPTTKPAATKPEMTTAKDKTTERDLTTPETTTAAPKMTKETATTEKTTESK 861
Db 618 -----TSPKT----- 622
QY 862 ITATTQVTSTTTQDTPPTTKLITLTKTTLAPKVTITTKITTTIMNKPEETAKPKDRAT 921
Db 623 -----TTLKATTLAPKVTAPE-----EIQNKPEPTTPASEDS 656
QY 922 NSKAT-----TPKP-QKPTKAPKPTSTTKPKTMPVRVKPKTPTPRKMTSTMPEN 972
Db 657 DSKTTLKPQKPTKAPKPTKPKTKAPKPKPTSTKKPKT-PKTRKPKTTAPLKTTSAPEN 715
QY 973 PFSRIAEAMLTTRNQTPNSKLVENVPKSDEAGAGETPHMLLRPHVFMPEVTPDMD 1032
Db 716 TTP--LEVMPLPTTTPKQTPNPETARVNDPDHEDADGGEKP-LIPGPVLPFTAIPGTD 772
QY 1033 YLPRVNOGIILINPMLS 1049
Db 773 LLAGRLNRGININPMS 789

RESULT 6
MUC2 HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Mucin 2 precursor (Intestinal mucin 2).
GN Name=MUC2; Synonyms=SMUC;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=94132002; PubMed=8300571;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;
RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.
RT Identification of the amino terminus and overall sequence similarity
RT to prepro-von Willebrand factor".
RL J. Biol. Chem. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=93016075; PubMed=1400449;
RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Rothe E.-M., Lagace R.E.,
RA Kim Y.S.;
RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located
RT both upstream and downstream of its central repetitive region.";
RL J. Biol. Chem. 267:21375-21383(1992).
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE=91358717; PubMed=1885763;
RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,
RA Petersen G.M., Kim Y.S.;
RT "MUC-2 human small intestinal mucin gene structure. Repeated arrays
RT and polymorphism.";
RL J. Clin. Invest. 88:1005-1013(1991).
CC -!- FUNCTION: Coats the epithelia of the intestines, airways, and
CC other mucus membrane-containing organs. Thought to provide a
CC protective, lubricating barrier against particles and infectious
CC agents at mucosal surfaces.
CC -!- SUBUNIT: Multimeric.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Colon, small intestine, colonic tumors,

```









```
QY 442 --TTKEPA-----PTTKSA-----PTTPKEPSTTKPEA- 470
D5 675 HETTAEPATKKTTHPTTKSTLRLITBEPTTKSSSTAKTTREPTTKKTERITERTTKEPTT 734
QY 471 -----PTTPK-----BPA-----PTTPKPPAPPTP 490
D5 735 RKTTHKTEBPTTKKTHPTTKSTLTKPTEPTTKSTTKTTRPTTKKTKTERTT 794
QY 491 KEPA-----PTTPKEPAPTTKKPA--PTAPK-----BPA----- 518
D5 795 KEPTTKTKTHKTEBPTNK-----TTTKKTHPTTKKSTLTKPTEPTTKSTTKTT 850
QY 519 -----PTTPKETA-----PTTPKLT-----PTT-PEK 540
D5 851 REPTTKRETTERITQEPSTKTTTHETAPATKKTTHPTTKSTLRLITBEPTTKSS 910
QY 541 LAPPTPEKAPPTPEELAPPTPEPT-----PTTPE--EPA----- 574
D5 911 TAKTTRE--PTTKRETTERTTKEPTTKTTHKTEBPTTKTTHPTTKKSTLTKPTE 967
QY 575 -PTTPKAAPNPKPAPTTPKEPAPTTPKEPA-----PTTPKETA-----P 615
D5 968 EPTTKSTTKTTR--PTTKKKTERTTKEPTTKTTHKTEBPTTKKTTTKEPTTKTTHP 1025
QY 616 TTPKETA-----PTTPKEPAPTTPKPAKELA-----PTTKKEPTSTSDKPA-- 659
D5 1026 TTKKSTLTKPTEPTTKSTTKTTRPTTKKVTERTTRETPTTKSTTKTTHETAPATKK 1085
QY 660 -----PTTPKETA-----PTTPKEPAPTTPKEPA-----PTTPKETAAPTTLKE 697
D5 1086 TTHPTTKSTLRLITBEPTTKSSSTAKTTREPTTKKRETTERTTKEPTTKTTHKTTT 1145
QY 698 PA-----PTTPK-----KPAKELAPTTTKGPTSTSDKPAAPTTPKETAAPTTPKEPA-- 744
D5 1146 PTTKKTHPTTKKSTLTKPTEB--PTTRK--TSTTKTTRPTTKKKTERTTKEPTTR 1200
QY 745 -----PTTPK-----KP-----APTTPPTPTTSEVST 768
D5 1201 KTTTHKTEBPTTKTTHPTTKKSTLTKPTEBPTTKSTTKTTRPTTKKTTTKEPTT 1260
QY 769 PTTPKEP-----TTPHKSDE-----STPELSAEPKALENSPEKPGVPTTKTAAKXP 818
D5 1261 ERTRPTTKKTHKTEBPTTKKTTTKEPTTKK--STLTKTEBPTTKKTTSTTKT 1319
QY 819 EMTTAKDKTTERDLR-----TTPETTTAAP----- 844
D5 1320 TREPTTKRVTTERTTRETPTTKTTHKTEBPTTKKTTTKEPTTKSTTKLTKPTEP 1379
QY 845 -----KMTKETAT--TTEKTESKITATTQVTSSTTQDTPPKITTLKT----- 887
D5 1380 TTKSTTKTTRPTTKRVTTERTTRETPTTKT--THKTEBPTTKKTTKTTKTTHEPTT 1437
QY 888 --TTLAP-----KVTTK-----KTIITTEIMNKP-----BEAKPKDRATN 922
D5 1438 KKSSTLTKPTEBPTTKSTTKTTRPTTKKVTERTTRETPTTKTTHKTEBPTTKKT 1497
QY 923 SKATTPKPO-----KP-----TKAPKKPT----- 941
D5 1498 TKKTHPTTKKSTLTKPTEBPTTKSTTKATREPTTKRVTTERTTRETPTTKTTHAKT 1557
QY 942 -----STKKPKTPRVKPKKTTPTPRKWTSMPNELNPTSIAEAMLQ 983
D5 1558 TEPTTKKTTKTTKTHPTTKKSTTL-----KTEBPTTKSTTKTTRPTTKTSTSVK 1613
QY 984 TT-----TRENOTPNSKLVEVNPKSEDAGAGET----- 1013
D5 1614 TTADQITKRTTAEMSTTNQEPSTVETTTNNSQNSNTTSTTEBQHVHHHHHHVHKP 1673
QY 1014 ----PHMLLRPH----- 1021
D5 1674 ADLGSILPLPLDPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPL 1733
```

```
QY 1022 VFMEPVT-PDMDYLPVFN 1039
D5 1734 ISLSEISLNPPLPLPLPN 1752

RESULT 9
AAS64673 PRELIMINARY; PRT: 1761 AA.
AC AAS64673;
DT 01-APR-2004 (TRENBLrel. 27, Created)
DT 01-APR-2004 (TRENBLrel. 27, Last sequence update)
DT 01-APR-2004 (TRENBLrel. 27, Last annotation update)
DE CG33300-PA (Fragment).
GN CG33300.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Fabois B., Delecher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodagef, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
SEQUENCE FROM N.A.
RP MEDLINE=22426065; PubMed=12537568;
RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pauley J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Tabor P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).
RN [3]
```





```

Db      1558 TEEPTTKTKTKTHBETTKSTL-----KTEPTTRKTSYKTRPTTRETSTSVK 1613
QY      984 TT-----TRRQTFNSKLVEYNPKSEDAGAGET-----1013
Db      1614 TTAQTKTKTAEMSTTQEPSTVETTTNSSQSNNTTTESTTTEBQVHHHHHHIYHKP 1673
QY      1014 ----PHMLLRH-----1021
Db      1674 ADLGSILFEDLPLPLFLPWPPLPELPLPLPTALPPLPLPLPEVNLTA 1733
QY      1022 VFMEVET-PMDDYLPVRVN 1039
Db      1734 ISLPEISLPLPLPLEN 1752

RESULT 10
Q7PMD5
ID Q7PMD5 PRELIMINARY; PRT; 3150 AA.
AC Q7PMD5;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE ENSANGP0000004655 (Fragment).
GN Name=ENSANGG0000003651.
OS Anopheles gambiae str. FES1.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008980; EAA13969.2; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR006770; OGF_recept.
DR Pfam; PF04680; OGF_r_III; 80.
FT NON_TER 1
FT NON_TER 3150
SQ SEQUENCE 3150 AA; 322879 MW; 3C7B3D441CE8C839 CRC64;

Query Match 22.3%; Score 1243.6; DB 2; Length 3150;
Best Local Similarity 19.9%; Pred. No. 2.2e-16;
Matches 402; Conservative 111; Mismatches 397; Indels 1114; Gaps 74;

QY 70 SENQESSSSSSSSSTTWIKSSKNSAANRELQKLVKONKNRTKKKTPKPPVD 129
Db 870 TESTDTMSSASSTPEPST-----TSGTTR-----TTTRPTPTDTMS 907
QY 130 EAGS-----GLDNGDFVVTPTDSTTQHN-----153
Db 908 SASSTPEPSTPGTTRTTPRPTSTESTDTMSSASSTPEPSTPGTTRTTPRPTSTESTD 967
QY 154 -----KVSTSPKITTAKPINP-----RPSLP 174
Db 968 TTMSSASSTPEPSTPGTTRTTPRPTSTESTDTMSSASSTPEPSTKPGTTRTTPRPTT 1027
QY 175 PNDSTKESVLTYNKETTVEKTTTNNKOTSDGKEKTTSA-----216
Db 1028 ESTDTMSSASSTPEPSTPGTTRTTPRPTSTESTDTMSSASSTPEPSTPGTTRTPT 1086
QY 217 --KETQSIEKTSAKLAP---TSKVLAKPTPKAET-----TTKGPAL 253
Db 1087 RPTSTESTDTMSSASSTPEPSTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTR 1146
QY 254 TTPKEPTP-----TTPKEPAS-----269

```

---

```

Db      1147 TTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTP 1206
QY      270 ----TTPKEP-----TPTTIKSAPTTP-----KEPA-- 291
Db      1207 GTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPST 1266
QY      292 PTTTKSAPTTP-----1306
Db      1267 PGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPE 1326
QY      307 PTTTKBPAPTTPKEBAPT-----TTPKEPA--PTTKSAPTTP-----341
Db      1327 PTTTPDTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSA 1386
QY      342 KEPAP-----TTPKKAPT-----TPKEBAPTTPKEPTTTPKEBAPT-----379
Db      1387 STPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDT 1446
QY      380 ----TKEPAP-----TTPKEBAPT-----APKKAP-----TTPKEBAP-- 409
Db      1447 MSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTES 1506
QY      410 ----TTPKEBAPTTPKEPSTTPKEPA-----PTTT 436
Db      1507 TDTMSSAYTPDPGIPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTST 1565
QY      437 KSAPTTP-----TKEPA--PTTIKSAPTTP-----KEPSTTPKEBAPTTPKE 476
Db      1566 ESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPR 1625
QY      477 PAP-----TTPKKAP-----TTPK 491
Db      1626 PTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTPT 1685
QY      492 EBAPT-----TPKEBAPTTPKPA-----PTAP-----KEPAPT 520
Db      1686 RPTPTDSTMSSMSSESTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPST 1745
QY      521 TPKETAPTTPKLTPT-----TPE-----KLAPT-----T 545
Db      1746 TTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASST 1805
QY      546 PE-----KPAPTTPBELAPTTPPEPTPT-----568
Db      1806 PEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSS 1865
QY      569 ----TPEE-----PAPTTPKAAAPNPKBAP-----591
Db      1866 SASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTD 1925
QY      592 ----TTPKEBAP-----TTPKEBAPTTPK 611
Db      1926 TTMSSASSTPDPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTSTEST 1985
QY      612 ETA-----PTTPKGTAPTTLKEBAP-----TTPKKAP 639
Db      1986 DTTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPTST 2045
QY      640 KELAPT-----TKBPTST-----TSDKAPTTPKGTAPTTPKEP-- 674
Db      2046 TESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRTTPRPT 2105
QY      675 ----APTTP-----KEPAPTTPKGTAPTTP 694
Db      2106 STESTDTMSSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSMSASSTPEPSTTPGTTRT 2165
QY      695 LKEBAP-----TTPKKAPAKELA-----PTTKGPTSTT 723
Db      2166 PTRPTPTDSTMSSMSASSTTPGTTRTTPRPTSTESTDTMSSASSTPEPSTTPGTTRT 2225
QY      724 SDKAPT-----TPKE-----TAPTTPKEBAP-----745
Db      2226 PTRPTPTDSTMSSMSASSTPEPSTTPGTTRTTPRPTSTESTDTMSSMSASSTPEPSTTP 2285

```





```
Db 621 -----PPSP 624
QY 236 VLAKPTPKAETTKGPAITTKPEPTTTKEPAST-----TPKEPTTTIKSAPTTPK 288
Db 625 APSPAPPS-----PAPSPQPPSPVPQPPSPVPPSPKPPSPAPPSVPVPPSPAPPSPA 678
QY 289 EPAPTTKSAPTTKKEPAPTTPKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 344
Db 679 PPSAPFPNAPPSAPPLPSPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAP 738
QY 345 APPTTKKAPPT-----TPKEPAPTTPKEPTTPKAPTTTKKAPTTTKKAPTTTKKAP 390
Db 739 APPSPSPAPPSDEPPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPS 798
QY 391 PAPT-----APKPAPTTPKKEPAPTTPKAPTT-----418
Db 799 PPAFPSPAPPSPEPPSPAPPSPPAPPSPPAPPSPPAPPSPPAPPSPPAPPSPPAPPS 858
QY 419 -TTKEPSTTKKEPAPTTPKSAPTTKKAPTTTKKAPTTTKKAPTTTKKAPTTTKKAP 477
Db 859 PAPPSPAPPSPEPPSPAPPSLEPPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAP 917
QY 478 AP-----TPKKEPAPTTPKEPAPTTPKAPTTTKKAPTTTKKAPTTTKKAP 515
Db 918 SPERSPEPPPPPPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPS 977
QY 516 -----EPAPTTPKETAPTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLTTPKLT 556
Db 978 PPSPPSPAPPSPPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPS 1036
QY 557 LAPT-----TPEPTTPPEPAPTTPKAAAPN-----TPKEPAPTTPK 595
Db 1037 TPPTSPPPPPPPSPAPPSAPPLPSPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPS 1096
QY 596 EPAPTTPK-----EPAPTTPKETAPTTPKGTAPTL-----626
Db 1097 PPSAPPSPEPPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPS 1156
QY 627 -----KEPAPTTPKPKAPKELAPT-----TKEPT 651
Db 1157 SPEVPPTPPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPS 1216
QY 652 STTSKAPATPKATPTTKKEPAPTTPKAPTTTKGAPTTLKBPATTPKPKAPKEL 711
Db 1217 SPEPPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPS 1271
QY 712 APT--TTKGTSTTSKAPATPKETAPTTPKEPAPTTPKAPTTTPPT--TSEVSTP 769
Db 1272 SPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPS 1328
QY 770 TTTKEPTTIHKSPDESTPELSAETPKALE-----NSPKKEPV- 807
Db 1329 AAALPPLPSPAPPLPVPSPAPPSAPPSAPPSAPPSAPPSAPPSAPPSAPPSAP 1388
QY 808 --PTTKTP-----AATKPE-----819
Db 1389 PPPPTTPPLAPLPDCTLLAQALLSIPDAANSVFVVSAGLTPSVAPSTPELLASFCT 1448
QY 820 -----MTTA-----824
Db 1449 VCSQUTATATSLVGGSSRRNTNGSSGNYNGSGDAAIQIPAGNSDTRGGSSSG 1508
QY 825 -----KDKTTERDL-----RTTP-----837
Db 1509 GSSSWGPGTETAEMAVDAVDQGTQQLSIGVYTRTVVDRTPPSVSGNVTLSANRI 1568
QY 838 -----ETTTAAPTWT-----847
Db 1569 KQPSAVGEASLNLGSKQAMLLTISFSEVPAPFAPASLIIVTGALVAEWAADKMTFY 1628
QY 848 -----KETATTTKTTESKIT-----863
Db 1629 VLAMTLPAELVATAGSSSSSGTSRSGNGNGTAAAAAAPPPTTGTGRRALQOAAA 1688
```

RESULT 13  
SLP1\_CLOTH STANDARD; PRT; 1664 AA.  
AC Q06852;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, last sequence update)  
DT 05-JUN-2004 (Rel. 44, last annotation update)  
DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer protein 1).  
DE Name=olpB;  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1515;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIB 10582;  
RX MEDLINE=93209931; PubMed=8456832;  
RA Fujino T., Beguin P., Aubert J.-P.;  
RT "Organization of a Clostridium thermocellum gene cluster encoding the cellulosomal scaffolding protein Cipa and a protein possibly involved in attachment of the cellulosome to the cell surface."  
RL J. Bacteriol. 175:1891-1899(1993).  
CC -!- SUBUNIT: Assembled into mono-layered crystalline arrays.  
CC -!- SUBCELLULAR LOCATION: Cell wall.  
CC -!- SIMILARITY: Contains 4 S-layer homology (SLH) domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



ID	QBWQ4	PRELIMINARY;	PRT; 1349 AA.
AC	QBWQ4;		
DT	01-MAR-2002	(T=EMBLrel. 20, Created)	
DT	01-MAR-2002	(T=EMBLrel. 20, Last sequence update)	
DT	01-OCT-2003	(T=EMBLrel. 25, Last annotation update)	
DE	Mucin 5 (Fragment).		
GN	Name=MUC5AC;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21426417; PubMed=11535137;		
RA	Escande F., Aubert J.P., Porchet N., Buisine M.P.;		
RT	"Human mucin gene MUC5AC: organization of its 5'-region and central repetitive region.";		
RL	Biochem. J. 358:763-772(2001).		
DR	EMBL; AJ298318; CAC83675.1; -.		
FT	NON_TER 1		
FT	NON_TER 1349 1349		
SQ	SEQUENCE 1349 AA; 135599 MW; 4DC3C1544F1E5BA CRC64;		
	Query Match	20.0%; Score 1113.7; DB 2; Length 1349;	
	Best Local Similarity	23.6%; Pred. No. 2.7e-14;	
	Matches 347; Conservative	78; Mismatches 371; Indels 673; Gaps 42;	
QY	68 SVSENGESSSSSSSSSTTIWKIKSSNSAANRELKKLVKONKORTKKTPPKPPV 127	:	: : : : :
DB	1 TTSTTSAPTSTTPTSGT-----	:	: : : : :
QY	128 VDEAGSLDNGFKVTPTDSTTHQNKYSTSPKITTAKPINRP-----SLPNSDTS 180	:	: : : : :
DB	21 -----TSTPQTSTTS-----ASTTSITSGPGTTPSPVPTTSTSAPTSTTS 62	:	: : : : :
QY	181 KETSLTVNKETTIVETKETTTNNKQTSIDGKEK-----TTSKETQSI 222	:	: : : : :
DB	63 AANTTSISAPTTSTTSAPTSTTSASASKTGLGTTTSPPTTSTTSPTSASTS 122	:	: : : : :
QY	223 EKTSADLAP-----TSKLIAKTPKAET-----TTKGPAIATTPKEPTTT----- 263	:	: : : : :
DB	123 SKTSGPGTTPSPVTTTIFA---PRITSASTSTTPGG-TTP-SPVPTTSTASVK 177	:	: : : : :
QY	264 -----	:	: : : : :
DB	178 TSTSHVISKTTHSQPVTRDCHLRCTWKFDVPDPSPGHGDKETYNNIIIRGEKICR 237	:	: : : : :
QY	264 -PKE-----	:	: : : : :
DB	238 RPEILTLQCRASHPEVSIEHIGVVQCSEBGLVCRNODQQGFMCINYEVRVLCE 297	:	: : : : :
QY	267 -----PASTTP-----KEPT-----PTTKSAPTTKPEAPT-- 294	:	: : : : :
DB	298 TPFGCVTSPVTPAPSTPSGRASPQTOSTSSQSKXTILVITSTSTPQTSTTSAPTTS 357	:	: : : : :
QY	295 TKSAPTTKPEAPTTPKEAPTTPKEAPTTPK-----SAPTTKPEPA 345	:	: : : : :
DB	358 TTSAPTTTSAPTSTTSTPQTSTTSAPTSTTSAPTSTTSAPTSTTSAPTSTTS 417	:	: : : : :
QY	346 PTTKKKPAPTTPKEAPTTPKEAPTTPKEAPT-----TTKEAPT-- 387	:	: : : : :
DB	418 PTSTTSAPTSTTSAPTSTSTPQTSTSAATSTTSSSGTTPSPVTTTSTASVSKTS 477	:	: : : : :
QY	388 -----	:	: : : : :
DB	478 TSHVSVSKTTHSQPVTRDCHPRCTWTWFDVPDPSPGHGDKETYNNIIIRGEKICRRP 537	:	: : : : :
QY	389 KE-----	:	: : : : :
DB	538 QELTRLQCRASHPEVSIEHLGVQVCSREBGLVCRNODQQGFMCINYEVRVLCCETP 597	:	: : : : :
QY	391 -----PAPTAPKKAPT-----TPKEAPTTPKE 414	:	: : : : :

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Blazewicz J., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazewicz J., Zhang Q., Chen L.X.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfankuch C., Baldwin D.,  
 RA Balwle R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fogler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Ceiniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
 RT *melanogaster* euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminiker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Ceiniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomics perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Mira S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Ceiniker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.

RG FLYBASE;  
 RL Submitted (SBP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RG FLYBASE;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003495; AAN09587.1; -;  
 DR FLYBASE; FBgn0052602; CG32602.  
 SQ SEQUENCE 3432 AA; 349492 MW; FE907AC75578657A CRC64;  
 Query Match 19.6%; Score 1090.2; DB 2; Length 3432;  
 Best Local Similarity 29.6%; Pred. No. 2.4e-13;  
 Matches 433; Conservative 60; Mismatches 403; Indels 568; Gaps 81;  
 QY 56 CPDYSEFCAEHSVSE-----NOESSSSSSSSSSSSSTIWKSSKNSAANRE 102  
 DB 58 CPSHPTTACCHRIITSPATPHCHPHATATPHNCGHDTSSSRAT----- 102  
 QY 103 LQKKLVKDNKKNRTKKPKPPVVDVAGSLDNGDFKVT--TPDTSITQHNKYSTSPK 160  
 DB 103 -----TPKPTLETTIGT-----TFKTTQITTRGPTTKPTEGTSQA 138  
 QY 161 ITTAKP-----INPRSLPPNSDTSKETSLSL-----TVNKETTVEKETTINKQT--STD 208  
 DB 139 PTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPT 198  
 QY 209 GKEKTSKAKET--QSIKTSKAK--DLAPTSKVLAKPT--PKAETTTKGPALITPKPT-- 260  
 DB 199 G-----TSAPKPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPTGTTAKPTTLKPT 254  
 QY 261 -PTTPKPEASTTPKBPPTTIK-----SAPTPKPEAPTTTKSAPTTKPEAPTTKPE 313  
 DB 255 KPTTLKPTGTTAK-----PTTLKPTGTSAPKPTTLKPTGTTAK--PTTLK-PTGTTAKP 308  
 QY 314 A-----PTTPKPEAPTTTKBPAPTT-----TKSAPTTKPEAPTTTPK-----KP 352  
 DB 309 TTLKPTGTSAPKPTTLKPTGTTAK-----PTTLKPTGTSAPKPTTLKPTGTTAKPTTLK 365  
 QY 353 APPTPKPEPA-----PTTPKBPPTTPKPEPA-----PTTPKPEPA----- 384  
 DB 366 TEGTSAPKPTTLKPTGTSAPKPTTLK-PTGTSAPKPTTLKPTGTSAPKPTTLKPTGTSAPK 424  
 QY 385 PTTPKPEAPAPK-----KPA-----PTTPKPEAPTTKPEAPTTKPEAPTTKPEAPTT 434  
 DB 425 PTTLKPTAGTTAKPTTLKPTGTSAPKPTTLKPTGTTAK--PTTLKPTGTTAK--PT 478  
 QY 435 TTK-----SAPTTKPEPA-----PTT-----TKSAPTTKBPSPPTTKPEPA----- 470  
 DB 479 TLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLK-PTGTTAKPTTLKPTG 537  
 QY 471 -----PTTPKPEAPTTTPK-----KPA-----PTTPKPEAPTTTPK----- 499  
 DB 538 TSAQPTTLKTKGTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTL 597  
 QY 500 EPAPTTTKKPA-----PTAPKPEAPTTKPEAPTTPKKLT-----PTT----- 537  
 DB 598 KPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTA 657  
 QY 538 -PEKLAPTTPEKAPTT-----PEKLAPT--TPEBPT--PTTPEBAPTTPK-- 579  
 DB 658 KPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKLT 717  
 QY 580 ---AAAPNTPKPEAPTTKPEAPTTPK-----EPAPTTKPEAPTTKPGTA----- 622  
 DB 718 EGTSAPKPTTLKPTGTTAK-----PTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTS 774  
 QY 623 -PTTLKBPAPTTPK-----KPA-----PKELAPT--TTKEPTSTTSKDA----- 659  
 DB 775 QPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPT 834  
 QY 660 -----PTTPKGTAPTTKPEAPTTKPEAPTTKPGTA-----PTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTA 891  
 DB 835 EGTTAKPTTLKPTGTTAK-----PTTLKPTGTTAKPTTLKPTGTSAPKPTTLKPTGTTA



```
QY 692 -PTTLKEPAPTPK-----KPA-----PRELAPT-----TTKGPT 720
Db |||||
QY 892 KPTTLKPTTEGTAKPTTLKPDGTAKPTTLKPTTEGTSAKPTTLKPDGTAKPTTLKPT 951
Db |||||
QY 721 STTSDKPA-----PTTPKETAPTTKEPAPTPK-----
Db |||||
QY 952 EGTSAKPTTLKPTTEGTAKPTTLKPTTEGTTAK-----PTTLKPTTEGTSAKPTTLKPTTEGTTA 1008
Db |||||
QY 750 -----KPA-----PTT-----PETPPPTTSEVSTPTTK-----
Db |||||
QY 1009 KPTTLKPTTEGTSAPPTTLKPTTEGTSAPPTTLKPTDGTAKPTTLKPTTEGTSAPPTTLKPT 1068
Db |||||
QY 774 -----EPTTIHKSDESPELSAPPT-----KALENSKPEPGVPTTKPAATKPEMTT 822
Db |||||
QY 1069 EGTTAKPTTL--KPTTEGT--SAOPTTLKPTTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGT 1123
Db |||||
QY 823 TAKDKTTERDLRITPTTT-----AAPKMTKETATTTKTESKITATTT-----
Db |||||
QY 1124 SAKPTTLKPTTEGTTAKPTTLKPTTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGTTAKPTTLK 1183
Db |||||
QY 868 -----QVTSITTTQDIT-----PFKITLTKTTTAP-----K 893
Db |||||
QY 1184 PTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGTSAP 1243
Db |||||
QY 894 VTTTKKTIITTEIMNKPEETAKPD-----RATNSKATTPKPOKTKAPKKPT 941
Db |||||
QY 1244 PTTLKPTTEGTTA--KP-TTLKPTDGTAKPTTLKPTTEGTSAPPTTLKPTTEGTTA--KPT 1297
Db |||||
QY 942 -----STKKPKTM-----PRVRKP-----KITPTPRKMT 965
Db |||||
QY 1298 TLKPTTEGTTAKPTTLKPTDGTAKPTTLKPTTEGTSAPPTTLKPTTEGTTAKPTTLKPTTEGT 1357
Db |||||
QY 966 STMP-ELNPTSRIAEAMLOTTTRPNOTPNPSKLVEVNPKSEDAGGARGETPHMLLRPHVEM 1024
Db |||||
QY 1358 SAKPTTLKPTD--GTTAKPTTLKPTTEGTSAPPTTLKPT-----TEGTT-----AKPTTLK 1404
Db |||||
QY 1025 PEVTEMDYLPV--PNQGIINP 1046
Db |||||
QY 1405 P--TEGTSAPPTTLKPTTEGTTAKP 1426
Db |||||
```

Search completed: October 13, 2004, 12:18:04  
Job time : 142.337 secs

**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 21.3671 Seconds  
(without alignments)  
4723.689 Million cell updates/sec

Title: SEQ1-D  
Perfect score: 5576  
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....DMDYLPVFNQGIINPMLS 1049

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79: \*  
1: Pir1: \*  
2: Pir2: \*  
3: Pir3: \*  
4: Pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1283.1	23.0	3020	2 A43932	mucin 2 precursor,
2	1128.7	20.2	1664	2 T18262	S-layer protein -
3	1002.8	18.0	1274	2 T16251	hypothetical prote
4	993.7	17.8	1489	2 T31108	cyst germination s
5	988	17.7	7962	2 T38346	elastic titin - hu
6	975.4	17.5	3570	2 T45025	mucin MUC5B, trach
7	974.8	17.5	2187	2 T30826	nascent polypeptid
8	918	16.5	1367	1 S48478	Glucan 1,4-alpha-g
9	871.4	15.6	1188	2 S49915	extensin-like prot
10	861.6	15.5	1229	2 T25697	hypothetical prote
11	860.6	15.4	3507	2 T34513	hypothetical prote
12	857.5	15.4	6642	2 T29757	protein UNC-89 - C
13	835.4	15.0	2897	2 E48666	cell proliferation
14	834.9	15.0	3256	2 A48666	proline-rich pepti
15	826.9	14.8	5762	2 A41819	high molecular mas
16	822.8	14.8	1151	2 T18535	mucin 1 precursor,
17	818.2	14.7	1344	1 A35175	tenascin-X - bovin
18	768.9	13.8	2232	2 T34434	hypothetical prote
19	768.2	13.8	4135	2 T42629	hypothetical prote
20	762.9	13.7	1832	2 T31113	mucin-like glycopr
21	756.1	13.6	2142	2 B35098	MHC class III hist
22	746.2	13.4	3942	2 T42730	Bassoon protein -
23	744.5	13.4	990	2 T51618	nucleolar phosphop
24	740.2	13.3	5262	2 T03454	ALR protein - huma
25	739.5	13.3	2225	2 T26063	hypothetical prote
26	733.9	13.2	761	2 C84672	hypothetical prote
27	733.5	13.2	924	2 S27923	Gene Lf3 protein -
28	733.4	13.2	4548	1 S00657	apoptosis(a) (EC
29	731.3	13.1	971	2 T19431	hypothetical prote

## ALIGNMENTS

## RESULT 1

A43932  
mucin 2 precursor, intestinal - human (fragments)  
N: Alternate names: mucin SMUC-41  
C: Species: Homo sapiens (man)  
C: Date: 10-Mar-1993 #sequence revision 12-Apr-1996 #text change 09-Jul-2004  
C: Accession: A49963; A45106; B45106; A43932; B33532; A61257; P00328; P00329  
R: Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A: Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the  
A: Reference number: A49963; MUID:94132002; PMID:8300571  
A: Accession: A49963  
A: Molecule type: mRNA  
A: Residues: 1-639 <GUI>  
A: Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:I21998  
R: Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A: Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr  
A: Reference number: A45106; MUID:93016075; PMID:1400449  
A: Accession: A45106  
A: Status: not compared with conceptual translation  
A: Molecule type: mRNA  
A: Residues: 626-1895 <GU2>  
A: Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396  
A: Note: sequence extracted from NCBI backbone (NCBI:P:116706)  
A: Accession: B45106  
A: Status: not compared with conceptual translation  
A: Molecule type: mRNA  
A: Residues: 2037-3020 <GU3>  
A: Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398  
A: Experimental source: colon  
R: Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M  
J. Clin. Invest. 88, 1005-1013, 1991  
A: Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp  
A: Reference number: A43932; MUID:91358717; PMID:1885763  
A: Accession: A43932  
A: Molecule type: DNA  
A: Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A: Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864  
A: Note: sequence inconsistent with the nucleotide translation  
A: Note: sequence extracted from NCBI backbone (NCBI:P:116706)  
R: Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A: Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden  
A: Reference number: A33532; MUID:89197956; PMID:2703501  
A: Accession: B33532  
A: Molecule type: mRNA  
A: Residues: 1916-2193 <GU4>  
A: Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874  
A: Experimental source: intestine  
R: Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

MHC class III hist  
MHC class III hist  
microtubule-associ  
Bassoon protein -  
hypothetical prote  
hypothetical prote  
hypothetical prote  
ascites siologlyco  
probable tenascin  
ALR protein - huma  
hypothetical prote  
All-1 protein +GTE  
cell proliferation  
mucin, tracheobron  
versican precursor  
hypothetical prote

J. Clin. Invest. 87, 77-82, 1991  
A>Title: Human bronchus and intestine express the same mucin gene.  
A/Reference number: A61257; MUID:91086481; PMID:1985113  
A/Accession: A61257  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 71,1925-1948,'TTS',1952-1954 <JAN>  
A/Experimental source: bronchus  
R/Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Wang, D.; Jones, C.; Forstner, R.; Biochem. Biophys. Res. Commun. 183, 821-828, 1992  
A/Title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-  
A/Reference number: P00328; MUID:92198477; PMID:1550588  
A/Accession: P00328  
A/Molecule type: mRNA  
A/Residues: 2328-2468 <XUG>  
A/Cross-references: GB:M86523  
A/Experimental source: small intestine  
A/Accession: P00329  
A/Molecule type: protein  
A/Residues: 2328-2342,'K',2344-2354 <XUG1>  
C/Genetics:  
A/Genes: GDB:MUC2  
A/Cross-references: GDB:120203; OMIM:158370  
A/Map position: lip15.5-lip15.5  
C/Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von  
C/Keywords: glycoprotein; intestine; tandem repeat  
F:2746-2834/Domain: von Willebrand factor type C repeat homology <WVC>  
  
Query Match 23.0%; Score 1283.1; DB 2; Length 3020;  
Best Local Similarity 21.1%; Pred. No. 2.7e-18;  
Matches 393; Conservative 98; Mismatches 482; Indels 889; Gaps 57;  
  
QY 3 WK-----TLPIYLLLSVFIQVSSQELSKRCFESF-----E 38  
DB 649 WREHVNKDVCPNSQVFLNYLTTCQTCRSLSEADSHLEGFAPVDGCGCPDHTFLDE 708  
QY 39 RGR-----ECDC-----DAQCKYDKC----- 55  
DB 709 KGRCPFLAKSCVHRGLYLEAGDVVVRQBERCVCRDRLHCRQLIGOSCTAPKIHMC 768  
QY 56 -----CPD----- 58  
DB 769 SNLTALATSKPRALSCQTLAAGYHYHTECVSGCPCDGLMDGRCVVEKECPCVHNDL 828  
QY 59 ----- 60  
DB 829 YSSGAKIKVDCNTCKRGRWVCTQAVCHCTSIYSGHYITPDGKYVDFDGHCSYAVQ 888  
QY 61 SFCAEHS-----VSNQSSSSSSSSSSSTIWKIKSSKNSANRELQKLVKDNKK 114  
DB 889 DYCGNSLSGSFIIITENVPGTGTGTCASKAIFMGR-----ELKLEDKHR 936  
QY 115 -----NRT-----KKKPTPKPPVVDAGSGL 135  
DB 937 VTIQDEGHVAVYTRVGGVVLVVESTGLIIVDXKETTTFIKLAPSYKGTV-----CGL 991  
QY 136 -----DNGDFKV-----TPTDTSTQ----- 151  
DB 992 CGNFDHRSNNDFTTRDHMVVSSELDFGNSWEKAPTCDVSTNPPCSINPHRRSWARKQC 1051  
QY 152 -----HNKVSTSP----- 159  
DB 1052 SILKSSVFSICHSKVDPPFYACVHDSCTDGDCECFCSAVASVAOECTKGACVFW 1111  
QY 160 -----KITTAKPIN----- 168  
DB 1112 RTPDLCPICFYDYNPPHCEWHEPCGRNSETCETCTINGIHSNISVYLECPCPCXDR 1171  
QY 169 -----PRPSLP----- 174  
DB 1172 PIYEEDLKKVTDKCGCVEDTHYPPGASVPTETCKSCVCTNSSQVVCREEGKIINQ 1231  
QY 175 -----PNSDTSKE-----TSLTVNKET-----TVETKETTITNKQSTSD 208

DB 1232 TDGAFCYWEICGPNGTVEKHFNICSTITRPSLTITTTTLTPTTSFTTTTTPTTS 1291  
QY 209 GKEKTTISAK-----ETOSIEKTSAD----- 229  
DB 1292 STVLSTTPKLCCLWSMDWINEHPSGSDGDRPFDDGVCAPEDIECRSVKDPHLSLEQH 1351  
QY 230 -----LAPTSKVLAKPTPKAE 245  
DB 1352 GQKVCQDSVGFICKNEDQFGNGPFLCYDYKIRVNCWMDKICITTPSPPTTTPSPPT 1411  
QY 246 TTTKGPALTTTKEPTTTPKEPASTTTPKEPPTTIKSAP-TTPKEPAPTTTTSKAPTTPKE 304  
DB 1412 TTTTLPPTTTPSPPTTTTTPPTTTPSPPTTTTTLPLTTPTPSPPTTTPPTTTPPTTTPS 1471  
QY 305 PAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAP-TTP-----KEPAPTTPKKAPT 355  
DB 1472 P-TTTPPTTTPSPPT-TTTPPTTTPSPPTTTPSPPTTTPPTTTPPTTTPSPPTTTP 1529  
QY 356 TPKEPAPTTPKEPTTTPKEPAPTTPKEPAPTTPKEPAPTAPKAPKAPTTPKEPAPTTPKEP 415  
DB 1530 TTTTPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPPTTTPPTTTPSPPTTTPSPPTTTP 1589  
QY 416 APPTTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 475  
DB 1590 TTTTTPPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1641  
QY 476 EPAPTTPKKAP-TTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 534  
DB 1642 SPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1695  
QY 535 PTTPEKLAAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEP 590  
DB 1696 WTPS-PPTTPSSPITTTTPSSSITTP-SPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1745  
QY 591 PPTTPKEPAPTTPKEPAPTTP-----KETAPTTP-----KG----- 620  
DB 1746 TLPPTTTSPLTTPPLPSITPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1805  
QY 621 ----- 620  
DB 1806 TELIGVCGGWAANISCRATWYDPVPIGQLGQTVVCDVSGLICKNEDQKPGVPMAP 1865  
QY 621 -----TAPTTPKEPAPTTPKAPKAPKELAPT-TKETSTSTSDK-----AP 660  
DB 1866 CLNVEINVQCECVTQPTM-----TTTTTENPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1922  
QY 661 T-TPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPTTPKGTAPT 719  
DB 1923 TPTPTSTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1981  
QY 720 TST-TSDKAPTTPKAPTTPKEPAPTTPKEPAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 771  
DB 1982 TPTPTSTKSTVTPITTTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 2041  
QY 772 TK-----EPTTIHKSDESPKALNSPKALNSPKALNSPKALNSPKALNSPKALNSPKALNSPK 826  
DB 2042 TSTAPIAELTSSNPPSSSTPQTSRSTSPSTTLLSTLTPALEMTSTAPSTPTAPT 2101  
QY 827 KTTTERTDRTTPTTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTPKAPTTP 884  
DB 2102 TTSGGHTLSPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 2159  
QY 885 -LKTITLAPKVTTPKTTITTEIMNKPEE-----TA 914  
DB 2160 IIRITGLRPPSSVLICCVLNDTYVAFGEVYNGYDTCYFVNCSLSTLEFYNWSGPS 2219  
QY 915 KPKDRATNSKATTPKAPKPTKAPKPTSTKPKTPMVRKPKTTPTRKMTSTWPELNPT 974  
DB 2220 TPTPTPTPSK-STPTPSKPSSTPSKPTPGTKPPECPDFDPPR-----QENET 2265  
QY 975 SRIAEAMLOTTTRNQITNSKLVEVNP-----KSEDAGGA----- 1009

Db 2266 WMLCCEM-ATCKYNNITVEIIVKCEBPPMPTCSNGLQVRVEDPDGCGCWHCEDCYCTG 2324  
Qy 1010 -----EGETHMLLRPHVPEVTPDMD----- 1032  
Db 2325 WGDPHYVTFDGLYYSQNCCTYLVLE-----EISPSVDNFGVYDYNHCDNDKVCSPR 2378  
Qy 1033 -----YLPVFNQGI 1043  
Db 2379 TLIVRHETQEVLIKTVHMPQOVQVNRQAVLPYKKYGLVYQSGINYYVVDIPELGVL 2438  
Qy 1044 IN 1045  
Db 2439 VS 2440  
  
RESULT 2  
T18262  
S-layer protein - Clostridium thermocellum  
C:Species: Clostridium thermocellum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18262  
R:Fujino, T.; Beguin, P.; Aubert, J.P.  
J. Bacteriol. 175, 1891-1899, 1993  
A:Title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose  
e.  
A:Reference number: Z18847; MUID:93209931; PMID:8458832  
A:Accession: T18262  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1664 <FUS>  
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841  
  
Query Match  
Best Local Similarity 20.2%; Score 1128.7; DB 2; Length 1664;  
Matches 366; Conservative 126; Mismatches 418; Indels 783; Gaps 67;  
  
Qy 4 KTLPIVLLLLSVFV-----IQVSSQE 26  
Db 6 KVLISLLTLLLIISTTSVNMSPAEATPSIEMVLDKTEVHGVITATIKVNIKRLAGYQ 65  
Qy 27 LSKC-----GRCFSEFREGRECDCAQCKKDKCP----- 57  
Db 66 LNIKFDPEVLQVDPATGEEFT-----DKSNPVNRVLLTNSKYGPTPV 108  
Qy 58 -----DYESFCAEHS-----VSENOES-----SSSS 80  
Db 109 AGNDIKSGIINFATGYNLLTAYKSSGIDHTGIIIGEIGFVKLKKQNTSIRFEDTLSPGA 168  
Qy 81 SSSSSSTIW-----KIKSSKNSAANRELQK-KLKVKD-----N 112  
Db 169 ISGTSLFWDABITFGYEVIOPLIWEAEPLKDSVALELDKTKVKVGDIITAIKIKEN 228  
Qy 113 KKN-----RUKKKT-----PKPPVDEAGSG 134  
Db 229 MKNFAGYQINIKYDPTMLEALETGSIAKRTWPVGTGLQSDNYGKTTAVANDVGAG 288  
Qy 135 LDN-----GDFKVTTPD----- 147  
Db 289 IINFARAYSNLTKYRETGVABETGIIKIGFVLKAGSTAIRFEDTAMPGAIGCTMYFD 348  
Qy 148 -----STQHNKV-----STSPKITTAKPINRPSLP-----PNS--- 177  
Db 349 WYGENIKGYSVQVQGEIVAEGERPEETEPVPTETVDTPTVTEEPVPSLPDSVI 408  
Qy 178 -----DTSK-----EFTSLVNKETTVEVK 196  
Db 409 MELDKTKVKVGDIITAIKIKENKNFAGYQINIKYDPTMLEALETGSIAKRTWPVGT 468  
Qy 197 ETTTNKQSTDGKERTTS-----AKETOSIEKTSKD 229  
Db 469 GTV-----LQSDNYGKTTAVANDVGAGIINFARAYSNLTKYRETGVABETGIIKIGFRV 523  
Qy 230 LAPTSKVLAKPTPKAETTT-----KGFALTTP-----KEP 259

Db 524 LKAGSTAI-----RFEDTTAMPGAIEGTYMFDWYGENIKGYSVQPGBIVAEGEETPEP 578  
Qy 260 TPT-TPKEPASTTKPEPTPT----- 278  
Db 579 VPTETPVDPPTVTEEPVPSLPSYIMELDKTKVKEGDVIIATIRVNNIKNLAGYQIG 638  
Qy 279 ----- 278  
Db 639 IKYDPKVLAEAFNIETGDPIDEGTWPVAVGGTILKNRDYLPCTGVAINNVSKGLINFAAYVY 698  
Qy 279 -----TIKSAPTPK----- 288  
Db 699 FDDYREEGKSEDTGIIIGNIGFVLKABDTTIRFELESMPGSDGTWMLDYLWLNRIISGVY 758  
Qy 289 --EPAPTTTKSAPTTKPEPAPTTTKEPAPTTTKEPAP-----TTTKEPAPTTKSAPTTPK 342  
Db 759 VIQAPAPIKAAS-----DEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 813  
Qy 343 EPAPTTKPKAPAPTTKPEPAP-----TPK-EPTPT-TPKEPAP-TKEPAPTTTKE 390  
Db 814 EPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 873  
Qy 391 PAPTAPKKA--PPTTKEPAP-TKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 447  
Db 874 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 930  
Qy 448 PTTTTSAPTTKPEPSPTTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 493  
Db 931 PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 990  
Qy 494 APT-TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 552  
Db 991 TSETPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1030  
Qy 553 ----TPEELAPTTPEEPTPT-TPEEAPTTTPEEAPTTTPEEAPTTTPEEAPTTTPEEAP 606  
Db 1031 DEPTPSD-EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1089  
Qy 607 PT-TPKETAPTTPKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 664  
Db 1090 PSETPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1146  
Qy 665 GTAPTTKPEPAPTTKPEPAPTTKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 723  
Db 1147 SDEPTPSDEPTPS--DEPTPS--DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1195  
Qy 724 SDKPAPT--TPKETAPTTKPEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 779  
Db 1196 SDEPTPSDEPTPSD-EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1249  
Qy 780 KSPDESTPELSAETPKALENSPKPEGVPTTKTAAKPEMTTAKDKTTERDLRTTPEP 839  
Db 1250 -PSDEPTP--SDEPTP--SETPEEPIPTDPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1302  
Qy 840 TTAAPKMTKETAATTTKTESKITAATTVQVTSITTTQDTTPKTLTKTLTKTLAKVTTTK 899  
Db 1303 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1347  
Qy 900 TITTEIMNKPEETAKPKDRATNSKATTPKQ-KPKAKPKPTSTKPKKTPRVRKPKPT 958  
Db 1348 -----PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1384  
Qy 959 PTPRKMTS-----TPELNPSTRIAEAMLQTTTRPNQTPNSK 995  
Db 1385 PTPPTPTPTSGSGSGSGSGGGGGGGGGTVPPTPTPTPS-----KPTSTPAPT 1433  
Qy 996 LVEVNPKSDAGGAGETPHMLLRPH---VFMBE----- 1026  
Db 1434 EIE-EPTPSDVPGAIGGEHAYLRGVDPGSGFRPERNITRAEAAVIFAKLGADESQAS 1492  
Qy 1027 -----VTFD-----MDYLPVFNQ 1041

Db 1493 ASPYSLADTHAAWAIAKFAATSGQLFKYDPDGTGTPKPDONITRABFATVVLHFLTKVKQOE 1552

QY 1042 I-----IINP 1046

Db 1553 IMSKLATIDISNP 1565

RESULT 3

T16251

hypothetical protein F35A5.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C:Accession: T16251

R:Leimbach, D.

submitted to the EMBL Data Library, January 1996

A:Description: The sequence of C. elegans cosmid F35A5.

A:Reference number: Z18485

A:Accession: T16251

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1274 <LEI>

A:Cross-references: UNIPROT:Q20007; EMBL:U46675; NID:g1166613; PID:g1166621; PIDN:AAB526

A:Experimental source: strain Bristol N2; clone F35A5

C:Genetics:

A:Gene: CESP:F35A5.1

A:Map position: X

A:Introns: 1272/2

Query Match 18.0%; Score 1002.8; DB 2; Length 1274;

Best Local Similarity 23.1%; Pred. No. 3.2e-13;

Matches 330; Conservative 99; Mismatches 309; Indels 692; Gaps 75;

QY 117 TKKKTP-----KPP--VDEAGSLDNGDFKVTTPDTSTTQHNVSTSPKITTAKP 166

Db 2 SRAPPTPIKNPAKKWPPWESVD----- 25

QY 167 INRPSPNPSDTSKETSLSLVNKKETVETKETTNNKQSTDCGKETSATQSIKTS 226

Db 26 -----EEMEVDDETAPSK-----LEKKPSLR----- 49

QY 227 AKDLAPTSKVLAKPTPKA-----ETTTKGPAITTP-----KEPTP- 261

Db 50 -KD-APTKPV---PSPGAPSPVPIKNPVKKWAPWEDDEPWEAPAPVPAKKVRDPSK 104

QY 262 TTPKEPASTTPKE-----PTPT-----TTKSAPT 287

Db 105 KVAKPEDASPKKIMAAKBPETLPAVPPTPVKNPVKKFKAPWEDDDEVVDKDAPTVP 164

QY 288 -----KEPA-----PTTKSAPTTPKEP--APTTPKEP-----PTT 317

Db 165 AKKTPVLKKKEPAAAKPRDPSPKAAAPSKBHDPIVPTPIKNPAKKWKPWEDDEVPT 224

QY 318 P-KEPAPTTPKEP-----PTTKSAPT----- 339

Db 225 EIKPEPEATRKVPALKKKKEPSTSVKPSVSDPSFTKVPVKKEPVPPTPIKNPTKKWPPW 284

QY 340 ---TP-----KEPAP-----TTPKKPAPTTPKEPAPTTPK 366

Db 285 EDETPEVEVKPPVPEKKAPVLKKDDAPAAAKARDPSKAAPKKVFPSPVVP--PTPVK 343

QY 367 EPT-----PTTPKEPAPTTPK-----EPAPTTP-KEPAP----- 393

Db 344 NPVKYKPPWVDDEPABEVKKPSAPEKKTPLVKKEPEPSTTPSSDPSPKKAAPAVKP 403

QY 394 --TAPKAPAP-----TTPKEPAPTTP----- 412

Db 404 RDSGPKATPLQADPKAQEVPTPVKNPVKKYKPPWVDDEDPVEVKQPEAPAKKTPLV 463

QY 413 --KEPA-----PTTKKEP-----SPTTP-KEP----- 432

Db 464 KKKEPAAKDTAKFATSKTPTETPKKDPVKPRDPSPKVAAPKPDQAAPATPVKNPVKKWR 523

QY 433 -----PTTKSAPT--TTKEPAP-----TTTKSAPTTPKEPSPPTTKE 468

Db 524 PPWEDDETPADDVSKPTDAKKTPTSLAKDKDPAPAKESLKPADTK-APAKPRDPSP---XK 579

QY 469 PAPTTPKEPAPTTPKK---PAPTTPKEPAPTTPKEP-----APTTPKK 508

Db 580 VAPTAPKKTPLVLAKEPAGPADSKTKEPEKSKPRDPSKKAAPVPAKVPKTEVAPAAVKK 639

QY 509 PAP-----TAPKEPAPTTPKETAPTTP-----KKLTP-----TTEKLAP 543

Db 640 PEPISKPKDTAPKKAEPNSP--VVPPTPVKNPVKKWKPWEDDDAPAKFVSLEPEKKTTP 697

QY 544 TTPKEPAPTTP-----ELAPTTPKEPAPTTP-----TTEPEEAPTTP----- 577

Db 698 VLAKK-APTKEPDEAAADPVSGSPSSKDKLAKKAAVKPRDPSKKAAPVPAKTEVPPA 756

QY 578 -----PK--AAAPNTPEPAPTTPKEP----- 597

Db 757 VVKKEPVAKSRDPSFKKAKAEPNSFVVP--PTPVKNPVKKWKPWEDDDAPAEVNVVP 815

QY 598 -----APTTPKEPAPTTPKETAPTTPKGTAAPTTLKEPAPTTP--PKKAPAKP----- 641

Db 816 EKKTPVLAKKTPVKPRDPSKKAAPKAPKSTKTDAFPVSVKKEPVPKPEPSPKKAEPNS 875

QY 642 --LAPTTPKEPT-----STTSKAPAPTTPKGTAAPT--TPKEP-----APTTPKEP 682

Db 876 PVVPPTPVKNPVKKWKPWEDDDDEPTEEVKPSPEKKTPLVLAKEPEKPKDAPKVAAP 935

QY 683 APTTPKGTAAPTTLKEPAPTTP-----TPKKPAP-----KELAPTTPTKGTS-----TTS 724

Db 936 RDESPKKAEP--KEPAKVAAPRDLSPKKAIPANTQOEAAPTTPVKNPVKKWKPWEDD 993

QY 725 DKP-----APTTPKEP-----APTTPKEPAP-----TTPKKPAPTTPPETTPPTTSEVSTP 769

Db 994 DEPAEPVSAPEPEKKTPLVLAKEPAPKPRDPSKKAAPVAAKPDPKIPEVP-----P 1044

QY 770 TTTKEPTTHKSP-----DESTPELSA-EP---TPKALENSEKPEGV-----P 808

Db 1045 TPVKNPVKKWKPWEDDDDEPSEVSAPEPEKKTPLVLAKEPAPTTPKATKPDSEAAADPVSGP 1104

QY 809 TTKTBAAT-----KPEMTTAKDKTTERDLRTTPTT---TAAPKMTKTAATTEKTTS 860

Db 1105 TSKDPLSKKAPVEKPKPTDCKDKLKSPAKKPEKAPAPAAKPKWKPWDDDDPDEPEA 1164

QY 861 KITATTQTQVSTTTQDTPPKITTLKTTLTTLAPKVTTTTKTTITTEINKEEETAKPKDRA 920

Db 1165 DFTVPAPSKKPDDEDPADPLG-----GPKTKDPK-----LNKKAPAEKPKTEK- 1206

QY 921 TNSKATTPKQKPTKAPKKTSTTKPKTMPVRKPK-----TTPTPKMTSTM 968

Db 1207 -----PKPEVSKPEPKTEPPKP--AAPKKWKPWEDDDDEPDEADFTWPAPK----- 1253

QY 969 PELNPTSGIAEAMLQTTTRENQTPNSKLVEVNPKSED-----AGGAEGETP 1014

Db 1254 -----PTEDPADSLGGPKPKDP 1271

RESULT 4

T31108

cyst germination specific acidic repeat protein precursor - Phytophthora infestans

C:Species: Phytophthora infestans (potato late blight agent)

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T31108

R:Goernhardt, B.

submitted to the EMBL Data Library, April 1998

A:Reference number: Z20986

A:Accession: T31108

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1489 <GOE>

A:Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAO

C:Genetics:

A:Gene: car90

Query Match		17.8%;	Score 993.7;	DB 2;	Length 1489;
Best Local Similarity		26.3%;	Pred. No. 6e-13;	Indels 643;	Gaps 74;
Matches 400;		Conservative	57;	Mismatches 422;	
QY	27	LSCKGRCFESFGRGRC-	-----DC-----	DAQCK-----	KY 52
Db	43	IFCSGVGAEPV--GTACPKAGDVATSDCQPYLLSYNGAVCAVPVDAECALIHDDMGCEP			100
QY	53	DKCCPDYEFCABEHSVSENOESSSSSSSSSSSTTWIKSKSNKSANRELQKKLVKDN			112
Db	101	PK--TGTYTSAVEAETIAAANGESSGWTGH-			137
QY	113	KN--RTKKPKPKPPV--		VDEAGSGLDGDF--	140
Db	138	EEIIPARVNDVTDPIGVNCEVATETATQGHATGGKYDTPSTGTQGTGNTGNIHYG			197
QY	141	-----KV-----		TTPTDSTTT--	150
Db	198	STTTEGVTKGGYPTDAKVIDGETYLDYPTGITEIIEDGTPPGYGTTDGGTTTGGYTT			257
QY	151	-----QHNKVTSPKI-----		-----	161
Db	258	VDNTHETGAGGYDAGTBREYSTTGVGYSTBETEGQHVGGYGFSPDEAPTEGTYV			317
QY	162	-----TTAKPIN-----PRPSLP-----		-----	174
Db	318	PREETAAFSEDTYAPREVTPYAPTEKPYDVBEITYVTEESTYAPTKSETNAPTERMHY			377
QY	175	-----PNSDT-----		SKETSLVKNKETTVE--TKETT--TTNKQT	205
Db	378	AHIEKPCDTEVTVYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET			437
QY	206	STDGKSKTSKAKTOSIEKTSKADLAPTSKVLAKAPTKAETTTKGPAITTKPKERTPTT--			263
Db	438	TYAPTEETT-----YAPTEKTTYAPT-----			479
QY	264	-----PKE-----		PASTTPKEPTTIKSAPTTPKEPAPT	293
Db	480	APTKEITYAPTEETTYASTBETTYAPTEETTYAPTEETTYAPABETTYEPTEET--TYAPTEETTYAPT			538
QY	294	-TTKSAPTTPKEPAPT--TTKEPAPTTPKGPAPTTPKAPPTTKSAPT-----		-----T	340
Db	539	EEITYAPTEETTYAPTEETTYAPABETTYEPTEETTYAPTEETTYAPTEETTYAPTEET		YAPTEETMYAPTEET	597
QY	341	PKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPTTPKGPAPT		-----AP-----TTKEP	383
Db	598	TYAPTEETTYAPABETTYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP			657
QY	384	APTTPKEP-----APTAPKKPAPTTPKEPAPTTPKEPAPT--TTKEPSPPTTPKEP----			431
Db	658	ABETTYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPABETTYEPTEET			717
QY	432	--APT--TTKSAPT-----TTKEP-----APT--TTKSAPT--TPKEPSPPTTPKEP			469
Db	718	TYAPTEETTYAPTEETMYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP			777
QY	470	-----APTTPKEPAPTTPKGPAPTTPKGPAPT--TPKEPAPTTPKGPAPTAPKEP----			517
Db	778	TGETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET			837
QY	518	-----APTTPKETAPTTPKLPPTTPKEKLAPT--TPKEP-----APTTPKEKLAPT			559
Db	838	TYTPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP			897
QY	560	T-----APTTPKETAPTTPKLPPTTPKEKLAPT--TPKEP-----APTTPKEKLAPT			587
Db	898	TEETTYASTBETTYAPTEETTYAPABETTYEPTEETTYAPTEETTYAPTEETTYAPTEET			957
QY	588	EPAPT--TPKEPAPTTPKEP-----APTTPKETAPTTPKGPAPT--			629
Db	958	TYAPTEETTYAPABETTYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP			1017
QY	630	APTTPKEP-----APKE-----LAPT-----		TTKEPTSTTSDKPA-----PITPK	664

Db	1018	ABETTYEPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET			1077
QY	665	GTAPTPKEPAPTTPKGPAPTTPKGPAPT--TLKEPAPTTPKGPAPT-----APKE-----LAP			713
Db	1078	TVAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAP			1137
QY	714	T-----TTKGPTSTSDKPAPTTPKEAPT--TPKEPAPTTPKGPAPT-----APT			755
Db	1138	TEETMYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTE			1195
QY	756	PETPPPTTSEVSTPT--TTKEPTTIHKSPDESTPELSAETPKALENSPKEPGVPTTKTP			813
Db	1196	EITYAPTEETTYAPTEETTYAPT--EETTYAPTEETTYAPT--EETTYAPTEETTYAPT			1249
QY	814	AATKPEMTTAKDKTTERDLRTPPTTAAPKMKET-----ATT-----TEKTIESKIUA			864
Db	1250	-----TEETTYAPTEETTYAPT--TEETMYAPTEETTYAPTEETTYAPTEA			1292
QY	865	TTTQVNTSTT--TODTT--PPKITT--LKTITLAPKVVTT-----KKT			900
Db	1293	TYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEETTYAPTEET			1352
QY	901	ITTEIMNKPEETAKPKDRATNSKATTPKPQ-----KPKAPKPTSTKKPKMPRVKRP--			955
Db	1353	VSTKEPCNTEETDEPTDEPTDPSDEPTDEPTDPLP-----TDEPST-----PCD			1401
QY	956	-----KTTPTPRK-----		-----MTSTMPEL	971
Db	1402	NOGINGIVENKVRYNAGIYNTTGPFRNSQSWHSCCRSCYNDPICHAFSFGHOTS-----			1456
QY	972	NPTSRIAEAMLQTTTRPNOFEN 993			
Db	1457	--SDSVCELTTSTSDREDOQN 1476			
RESULT 5					
I38346					
elastic titin - human (fragment)					
C;Species: Homo sapiens (man)					
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004					
C;Accession: I38346					
R;Label: S.; Kolmerer, B.					
Science 270, 293-296, 1995					
A;Title: titins: giant proteins in charge of muscle ultrastructure and elasticity.					
A;Reference number: A57430; MUID:96026330; PMID:7569978					
A;Accession: I38346					
A;Status: preliminary;					
A;Molecule type: mRNA					
A;Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g10:					
C;Genetics:					
A;Gene: GDB:TTN					
A;Cross-references: GDB:127867; OMIM:188840					
A;Map position: 2q31-2q31					
Query Match					
Best Local Similarity					
Matches 412;					
Conservative 116;					
Mismatches 410;					
Indels 2420;					
Gaps 99;					
QY	14	LSVFVIQOVSSQELSC-----KGRCFE-----			41
Db	4497	LEIFDVDTSQSGEYTCIVSNEAGKASCTTHLYIKAPAKFVKRLNDYSIEKGKPLILEGTF			4556
QY	42	-----ECD-----		-----C	45
Db	4557	TGTPPISVTKKNGINVTSPQRGNITTEKSPILEIPSTVEDAGQVNCYIENASGKDSC			4616
QY	46	DAQCKKYDKCCPY-----			59
Db	4617	SAQILILE---PPYFVKQLEPFPKVSVDGSLQCLAGTPEIGVSWYKGTUKLRPTTYK			4673
QY	60	-----ESFCABEHSVSENOESS-----			76

Db	4674	MHFRNNVATLVFNQVDINDSGEYICKAENSUGEVSASTFLTVQEQKLPPSFRQLRDVQE	4733
Qy	77	-----SSS-----	79
Db	4734	TVGLPVVDFCAISGSEPIVSWSYKDGKPLKDPNVQTSFLDNTATLNIKFTDRSLAGQYS	4793
Qy	80	-----SSSSSSSTI-----	88
Db	4794	CTATNPISASSARLLITTEKNPPFPDIRLAPVDVAVGESADFECHVTGTQPIKVSNAK	4853
Qy	89	-----WKIKSSKNSA-----	112
Db	4854	DSREIRSGGKYIOISYLENSAHLTVLKVDKDGSGQYTCYAVNEVGKSDCTAQLNIKERLIP	4913
Qy	113	-----KNRTTKKBT-----	122
Db	4914	PSFTKRLSETVEETEGNSFKLEGRVAGSQPITVAMYKNNIEIQTSNCEITFKNNTLVLQ	4973
Qy	123	-----PKPPVVDB-----	130
Db	4974	VRKAGMNDAGLYTCVSDNAGSALCTSSIVIKEPKPPVDFDQHLTPVTVSEGEVQLSCH	5033
Qy	131	-----AG-----	147
Db	5034	VOGSEPIRIOWLKAGREIKPSDRCSFSGTAGVLELRDVAKADSGDVVCKASNVAGSDT	5093
Qy	148	STTQHNKVSPTKITTAKPINPRSLPNSD-----	192
Db	5094	-----TKSKVT-----	5140
Qy	193	-----VETKETTTTN-----	202
Db	5141	KVGDDPIPNVKNWTKWKRQLNOGRVFIHQKDEAKLEIRDTYKDSGLRVCVAFNHEGE	5200
Qy	203	-----KQSTDG-----	209
Db	5201	IESNVNLQVDERKKQKIEGLDRLAMLKKTPLILKKGAGEEBEIDIMELLLKNVDKPEYKYA	5260
Qy	210	-----KEKTTSAKETQSIK-----	224
Db	5261	RMGIDTFRGLLOAFELLKQSBEEHRLBIEIERSEDERDEKFEELVSVFIQORLSQTEP	5320
Qy	225	-----TSAKD-----	235
Db	5321	VTLLKQIENQVLKONDVAFIDIKINYPEIKLSWYKTEKLBPSOKFEISIDGRHCLR	5380
Qy	236	-----	235
Db	5381	VKNQQLKDOGNRYRLVCGPHIASAKLTVIEPAWBRHLQDVTLKBGQTCMTVQFVSPNVKS	5440
Qy	236	-----	235
Db	5441	EWFRNGRILKPOGRHKTVEBKHVHLTIADVRAEDQGYTKCYEDLETSAELRIEAPIQ	5500
Qy	236	-----	235
Db	5501	FTXRIQNIUVSEHQSAFCEVSFDDAIVTWYKGPTELITESQKYNFRNDGRCHYMTIHNV	5560
Qy	236	-----VLAKPTKAEHTTKGPAITTPKE-----	262
Db	5561	TPDDEGVYSVIARLEPRGEARSTAEIYLTTKBIKLELKPDPIDPSRVPIPTMIRAIVPE	5620
Qy	263	-----TPKE-----	276
Db	5621	EIPPVAVPVLPTPPEKKPPPKRIEVTKKAVKDAKKVAKPKENTPREIIVKPPPP	5680
Qy	277	PTTI-----	280
Db	5681	PTTLIPAKAPEIIDVSSKABEVKIMTITRKKEVQKEKAEVYKQAVHKEKRVFISSFEE	5740
Qy	281	-----	280
Db	5741	PYDELEVEPYTEPEQPYVEBPDDEYBEIKVEAKKEVHEEEDFEEOEYEREBEGYDE	5800
Qy	281	-----KSAPTPK-----	288
Db	5801	GEESWEEAYOEREVIQVQKEVYESHERKVPKAPVEKKAPPPKVIKPKVIEKTSRR	5860
Qy	289	-----	288
Db	5861	MBEKVQVTKVPEVSKKIVQKPSRTFVQBEVIEVKVPAVHTKQWVISEEKMFASHTEE	5920
Qy	289	-----EPAPTTTKSAPTTTKBPA-----	312
Db	5921	EVSVTVEVQKEIVTEEKIHVAVSKRVEPPP-----	5976
Qy	313	PAPTPKPEPAPTTTKBPAITTKSAPTTTKBPA-----	354
Db	5977	PAPKVPPEVP-----	6031
Qy	355	-----TTPKE-----	389
Db	6032	APPAKVEVQGVVTEEKIIVTQREESPPPAVPEIPKKKVPKEKVPVPRKEEVEVPPPK	6091
Qy	390	EPAPTAPEKBPAP-----	401
Db	6092	VPA-----	6149
Qy	402	-----	401
Db	6150	TRHEVSABEESYSSEEBEGVSI SVYREEREEREAEVTEYVMEPEEYVVEEKLHIIS	6209
Qy	402	-----TTPKE-----	423
Db	6210	KRVEABPAEVTREOEKKIVLKPIKPIAKIEBPPPAKVPPEAPKKIVPEKKVPAPVPEKBP	6269
Qy	424	SPTTPKEP-APTTPKSAPTTT-----	457
Db	6270	PPKVPPEKPPVPEKKVPVKVIMKEBPLPAKVTEKHMQITQBEKVLVAVTKKEAPKARV	6329
Qy	458	PKPSPTTYKE-----	479
Db	6330	PEEPKRAVPEKVLKLPKREBPPPAKVTFRKRVVKEEKVSIAPKREPOPIKEVTIME	6389
Qy	480	-----	479
Db	6390	EKERAYTLBEEAVSVOREBEEYEBEYDYKBFEBEYPTBEBYDQVEEVEEYEBEYEBHEE	6449
Qy	480	-----TTPKPPAPT-----	517
Db	6450	YITEBKPPIPVKVPPEBPPVTKPKAPPAKVLKKAAPVEKVPVPIPKKLKPPPPKVPPEPK	6509
Qy	518	-----APTTPKETA	526
Db	6510	KVFEKIKH:SI TKREKEQVTEPAAKVPMKPKRVVABEKVPVPRKEVAPPVVRVPEVPKELE	6569
Qy	527	P-----	549
Db	6570	PEEVAPEEVEVTHVEBYLVEEBEYIHEBEPFITEBEEVDPVVKV-PEVPRKVPVPEKK	6628
Qy	550	-----APTTPPELAPT-----	576
Db	6629	PVPVPKKEAPPAKPAKVPPEVPKKEKVPVLI PKKEKPPPAKVPPEVPKVPPEKVPVVPK	6688
Qy	577	-----TPKAAAPNTKPEPAPTTTPKEPAPT-----	627
Db	6689	KVEAPPAKVPPEKVPPEVP-EKKVPVPEAPKPKVABAPPAKVPPEVPKCL:PEEKKETPVPPKVP	6747
Qy	628	EPAPTTTPKK-----	636
Db	6748	APPPKVPKKEPVPVPPVALPQBEELVFEIEIVPEBEEVLPEBEEVLPEBEEVLPEBEEVL	6807
Qy	637	-----PAP-----	639
Db	6808	EBEEIPPEBEEVPEBEEVPEBEEFVPEBEEVLPEVKKVPVPPAPVPEIKKKVTEKKVPI	6867



```
QY 640 -KELAP-----TTTKEP----- 650
Db 6868 KKEEAPPAKVPVKKVEEKRIILPKKEBVLVPVEEPEEPEEPISEEEIPEPPPSIEEVE 6927
QY 651 -----TSTTSKAPATPKGTAPT-----PKE--PAPTTPKEPAP----- 684
Db 6928 EVAPPRVPEIKAVPEATPVPKVEAPPAKVKKIPEEKVPVQKKEAPPAKVPVP 6987
QY 685 ----- 684
Db 6988 KKVPKLVLPKEAVPAKGRTVLEKSVAPRQBVVVVKERLELEVAEVEEIPPEEE 7047
QY 685 -----TTPKG 689
Db 7048 FHEVEEYFEEGFEHEVEEFIKLEQHRVEBEHRVEKVRHVIEWFEABEVEFEKPKAPPKG 7107
QY 690 -----TAPTTTL--KEP--APTTPKK-----PA-----PK 709
Db 7108 PEISEKLIIPKPKPTKVVRKEPPAKVPVPPKLIIVVEEKVRVPEEPRVPTKVPVLPK 7167
QY 710 ELAPTTPKGTSTSDKAPTTPKETAPTTPKEPAPTTPKKAPATTPE-----TPPPTTSE 765
Db 7168 EVVP-----EKKVPVPAK-----KPEAPPKVPPEAPKEVVPVPEKVPVPEPKKPE 7212
QY 766 VSTPTTKE-----PTTIHKSDESTBELSAEP--TPKALENSKEPFGVPTT-- 810
Db 7213 V-PPTKVPEVPKAAVPEKKVPEAIPPKPEPPPEVEEPEEESPSA---PPKKDEVPVVRV 7268
QY 811 -----KTPAA--TKPEMTTAAKDKTTERDLRTTPTTTTAAAPKMTKETATTTEKT 857
Db 7269 PEVPKEVVPKVPVPAAPKKPEV-----TPVKVPEAPK-----EVV 7304
QY 858 TESKITATTQVSTTTQDTPPTPKITLTKTLTLAPKVTTKITTT-----EIM 907
Db 7305 PEKKV-----PVPPKPEVPPTKVPEV-----PKVAVPEKKVPEAIPPKPESPPEVF 7353
QY 908 NKPEETA----- 914
Db 7354 EEPPEVALEBPPEAEVSEPEPAAPPQVTPPKNPVPEKKAPAVVAKPELPPVKVPEVPK 7413
QY 915 -----KPKDRATNSKATTPK-----POKPTKAPKKPTSTK 944
Db 7414 EVVPEKKVPLVVPKPEAPPAKVPVPEVPEKKVAVPKPEVPAKVPVPEKKVLEE 7473
QY 945 KP----- 946
Db 7474 KPAVPVPERAESPPPEVVEPEEPIAPEEIEAPEEKVPVVAESEEPEVPPVPAVPEEPKI 7533
QY 947 -----KTMPRVRPKTTP----- 959
Db 7534 IPEKKVPVIRKPEAPPPEPEPEKVBKPKLKRPPPPPPAPPKEDVKEKIFOLKAIPKK 7593
QY 960 -----TPRKMTS----- 966
Db 7594 KVPENQVPEKVELTFLKVPGEKKVKRLIPERKPEKEEVLKVLKRPEEPEEKVPE 7653
QY 967 -----TWPELNPTSRJABAMLQTTTR-----PNQTPNSK-----LVEVNPKE 1004
Db 7654 KKLEKVKKPAVPEPPPKPEVEEVVPTVKRERKIPEPTKVPEIKPAIPLPAPEPKPE 7713
QY 1005 DAGGAEGETHMLLRPHVFPVEVTP-----DMDYLPVFNQGI 1042
Db 7714 -----AEVKT-----IKPPPEVPEPTPIAAPVTVVVGKKAERAKPEBAKPKGIKV 7763
```

## RESULT 6

T45025

mucin MUC5B, tracheobronchial [imported] - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000

C;Accession: T45025

R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.

J. Biol. Chem. 272, 3168-3178, 1997

A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat

A;Reference number: Z22899; MUID:97166151; PMID:9013550

A;Accession: T45025

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-3570 &lt;DES&gt;

A;Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503

A;Experimental source: placenta

C;Genetics:

A;Gene: MUC5B

```
Query Match 17.5%; Score 975.4; DB 2; Length 3570;
Best Local Similarity 13.2%; Pred. No. 4.6e-12;
Matches 433; Conservative 129; Mismatches 430; Indels 2286; Gaps 96;

QY 6 LPTYLALLLSVFIQOVSSQELSK-----GRCFESFER----- 39
Db 3 LPV-----STVCREV-----CRWSWNGHRPBPGLGGGDFETENLRQGYQVCPV 50
QY 40 -----GREDCD-----AQCKYDKCCPDYE--SFCAE----- 65
Db 51 LADIECRAALPMPLEBELGQQVDCDRMGLMCANSQQSPPLCHDYELRVLCCYVPCGP 110
QY 66 -----BHSVSENQ-----SSSSSSSSSSSTIWKIKSKNSAA----- 99
Db 111 SPAGTSPQPSLSASTEPAPVPTQTATEKTLWVTPSIRSTAALTQTGSSSGPVTVT 170
QY 100 ----- 99
Db 171 PSAPGTTTCQPCQMTWFDYDYPKSEQLGGDVESYDKIRAAAGHLCCQPKDIECOAESF 230
QY 100 -----NRELOKLVKDKNKQPT-----KKKTEKPEPV 128
Db 231 PNWTLAQGVQKCHDVHGLVCRNWEQGVKCYNYRIVLCCDDHCHGRATTPPTT 290
QY 129 DEAGSLDNGDFKVTPTDSTTC-----HNKVTSPKITTAKPINP----- 169
Db 291 E-----LETATYTTQALSTPQPTSPGLTRAPPASTAVPTLSGLTSPRY 338
QY 170 -----RPSLPN----- 176
Db 339 TSLGTATTGGPRQSGASTEPTVGVATSLPTRSALPGTTGSLGTWRPSQPTLAPT 398
QY 177 -----SDTSKE--TSL--TVNKETTVEKETTNTKQTSDDGKEKTSKET- 219
Db 399 ATSRARPTGTASTASKEPLTSLAPLTSELSTQAEISTPTRETTSPLNTTISQGT 458
QY 220 -----QSIK----- 223
Db 459 RCQPKCEWTEWFDVDPFTSGVASGDMETPENIRAAGKMCWAPKSIICRAENYPEVSDQ 518
QY 224 -----KTSANDLAP- 232
Db 519 VGQVLTCSLGTCLCKNDQTRFNMCFNYNVRVLCDDYSHCPSTLSTATSPSTPG 578
QY 233 TSKVLAKP-----TPKAETTTKGALTPKEPTP----- 261
Db 579 TFWILTPTTTATTASTGTASTASQATAGTHVSTTATTPTVTSSK-ATFPSSPGTA 637
QY 262 -----TTPKEPASTTPKEPTPTTIK----- 281
Db 638 TALPALRSTATTATTSFTAI PSSSLGTWTRLSQITTPMATWSTATSPSTPTVHTSV 697
QY 282 -----SAPT 286
Db 698 LTTTATTGATGVATPSSSTPGTAHTTKVLTITTTTGFTATPSSSPGRARLPLVMISTTT 757
QY 287 P-----KEPAPTTT----- 295
Db 758 PTTTRGSTVTPSSIPGTHHTPTVLTTTITTTTATGSMATPSSSTQTSGPPSLTITATITA 817
QY 296 -----KSAP-ITPKEPAPTTT-KEPAPT-----TPKEPAPTTTKEPAPTTTKS----- 336
```

```
Db      818  TGSNTNPSPTGTPPIPVLTATTAATPAATSSSTVTPSSALGTHTPVPVNTTATTHGRSL  877
QY      337  APTTP-----KEPAPTPKKPAPT-----PKEPAPTPKEPTPT  372
Db      878  SPSPHVCTAMTSATSGILGTHITPEPSTGTSHTPAATGTTQHSFALSSPHPSRRT  937
QY      373  PKEPAPTPKEPAPT--TPKBPAPTAPKKPAPT--PKEP--APT-----  411
Db      938  ESPSPGTTTGHGTATSRITATATPKTRTSILLPSOFTSAPITTVVMGCEPQCAWSE  997
QY      412  -----PKEPAPTTTKPS-----  411
Db      998  WLDYSYMPGPGSGDFDYSNIRAGGAVCEQPLGLECRAQAQGVPLRELQGVVECSLD  1057
QY      412  -----PKEPAPTTTKPS-----  424
Db      1058  FGLVCRNREQVKMCFNVEIRVFCNCGHCPSTPATSSPTATSSPTGTTWILTELT  1117
QY      425  -----PTTPKEPAP-----TTTKSAPT-----  441
Db      1118  ATTTESTGSTATFTSLRTAPPPKVLTTTATPTVTSKATPSSPGTATAPALRSTAT  1177
QY      442  -----  441
Db      1178  TPTATSVPIPPSSSLGTTWTRLSQTTTPTATMTATPSSPTETAHTSTVLTAATTTGAT  1237
QY      442  -----TTKEPAPTT-----KSAPTPKEP-----SPT-----  465
Db      1238  GSVATPSSPTGTAHTTKVPTTTTGTATPSSPGTALTTPVWISTTTTTTTRGSTVPS  1297
QY      466  -----  473
Db      1298  SIPGTHTAIVLTITTTTATGSMATPSSSTQTSQTPPSLTITTTATITATGSTNPSSTP  1357
QY      474  PKEPAP-----TPKBPAPT-----TPKEPAPTTPKEPAPT-----  506
Db      1358  GTRPIPVLTATTATPAATSSSTVTPSSALGTHTPVPVNTTATTHGRSLSPSPHTVTA  1417
QY      507  -----  522
Db      1418  WTSATSGTLGTHITPESGTSHTPAATGTTQHSPTALSSPHPSRTTESPSGTTTT  1477
QY      523  -----KETAPTPKK-----LTPTTPEKLAPTTP-----  546
Db      1478  GHTATSRITATATPKTRTSILLPSPTS-APITTVTMGCEPQCAWSEWLDYSYMPG  1536
QY      547  -----  549
Db      1537  PGGGDFDYSNIRAGGAVCEQPLGLECRAQAQGVPLRELQGVVECSLDGFLVCRNREQ  1596
QY      550  -----APTTPBELAPTTPEE-----  564
Db      1597  VGKFKMCFNVEIRVFCNCGHCPSTPATSSPTGTTWILTEQTTAAITATGST  1656
QY      565  -----PTPTTPEEPAP-----TTP-----KAAAPNTPK-----EPA  590
Db      1657  AIPSPGTTAPPPKVLTSQATTPATSSKATSSSPRTATTLPLVLTSTATKSTATSFTPI  1716
QY      591  PTT-----PKEP-----APTTPKEPAPTTPKETA  614
Db      1717  PSTSLGTTGTSQNRPHPMATWSTIHPSSTPETHSTSTVLTTKATTTTRATSSMSTP  1776
QY      615  PTT-----PKGAPT-----LKEPAPT-----  632
Db      1777  GITWILTELTAAATTAALPHGTPSPSTPGTTWILTEPSTAIIVTGTATASSTRATA  1836
QY      633  -----TPKK-----PAPKELAPT-----  645
Db      1837  GTLKVLSTATTPTVISSRATPSSPGTATAPALRSTATTTATPATSVAIPSSSLGTAWT  1896
QY      646  -----  932
2973  TPETHSTVLTKATTRATSTSTPSSPTGTTWILTELTAAITATGATPSTPSTPG  3032
```



Db 1733 TLPSLKEASVLSSTATSSGK-----DSHISPVSDACSGNGTIT--PQASEKLPSK 1779  
QY 940 -----PTSTKPKXTWPR--VRK 954  
Db 1780 KGTAFTTEMLAAPAPESALAITAPIOKSPGANGSASSPKCPDPSSKKDTKGLPSAVALA 1839  
QY 955 PKTPTPRKWTSTWPELNPSTSRIAEAMLQTTTRPNQ----- 990  
Db 1840 PQIVPVEK-----DISKAETLVSFAKSGDCCLHSPKPGVGSQVATPLAAFTSDK 1889  
QY 991 -----TPNSKLVEVNPKESEDAGGAEGETPHMLLRPHVEMPE 1026  
Db 1890 VPPEAVSASVAPKPAPAPASLTLPAPSPVAPLPKPKLLESAPGVLESPSKL-----P 1941  
QY 1027 VTPDMYLPV-----PNOGIILN-----PML 1048  
Db 1942 VPAEDELPLIPPEAVSGGEPFQILVNNMPAKPAGTAPAPASAKQFVL 1991  
RESULT 8  
S48478  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR019c  
C:Species: Saccharomyces cerevisiae  
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: S48478; A26877; E26877; S27281; J06123  
R:Rowley, K.  
submitted to the EMBL Data Library, October 1994  
A:Reference number: S48478  
A:Accession: S48478  
A:Molecule type: DNA  
A:Residues: 1-1367 <ROW>  
A:Cross-references: UNIPROT:P08640; GB:247047; EMBL:Z38061; NID:g603997; PID:g763364; GS  
E:Yamashita, I.; Nakamura, M.; Fukui, S.  
J. Bacteriol. 169, 2142-2149, 1987  
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.  
A:Reference number: A91831; MUID:87194600; PMID:3106330  
A:Accession: A26877  
A:Molecule type: DNA  
A:Residues: 1-242 <YAM>  
A:Cross-references: EMBL:M16164; NID:g172522; PIDN:AAA35014.1; PID:g172525  
A:Accession: B26877  
A:Molecule type: DNA  
A:Residues: 762-1331 <YA2>  
A:Cross-references: EMBL:M16165; NID:g172523; PIDN:AAA35015.1; PID:g172526  
E:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
FEBS Lett. 239, 179-184, 1988  
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar  
A:Reference number: S27281; MUID:89031230; PMID:3141213  
A:Accession: S27281  
A:Molecule type: DNA  
A:Residues: 1-31 <PAR>  
A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552  
E:Lambrechts, M.G.; Bauer, F.P.; Marmur, J.; Pretorius, I.S.  
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy  
A:Reference number: J06123; MUID:96323237; PMID:8710886  
A:Accession: J06123  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1367 <LAW>  
A:Cross-references: GB:U30626; NID:g1304386; PIDN:AAC49609.1; PID:g1304387  
C:Genetics:  
A:Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458  
A:Cross-references: MIFS:YIR019C; SGD:S0001458  
A:Map position: 9R  
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
F:5-21/Domain: transmembrane #status predicted <TM1>  
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match

16.5%; Score 918; DB 1; Length 1367;

Best Local Similarity 25.1%; Pred. No. 1.7e-11;  
Matches 337; Conservative 108; Mismatches 437; Indels 460; Gaps 64;  
QY 9 YLL-----LLLSVFVIQOVSSQELSKGRC-----FE----- 35  
Db 118 YLIDNPDTFTATEVYATQDVS-----CQVMENFOIQFEYLOGSAAQYASSWQWG 169  
QY 36 --SFEGRGEC-----DCAOQCKYDKCCPDYDFSCAEHHSVENQ 73  
Db 170 TTFDILSTGCNNYDNOGHSQTDFFGYWNIDCDNNC-----GGTKSSTTTSTGS 218  
QY 74 ESSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDNKKNTKKKPTPKPPVVDGAS 133  
Db 219 ESSSTTSSESSTTS----- 235  
QY 134 GLDNGDFKVTTPDTSTTOHKNKSTSPKITTAKINP-----RPSLPNNSDTSKETSL 185  
Db 236 -----STSESSITTSSTSSSTSSSTTAPATPTTCTKEKPTPTTTSCTKEKPT 286  
QY 186 TVNKETTVETKETTINKQTSIDGKKTTSKAKTOSIEKTSKADLAPTSKVLAKPTPAE 245  
Db 287 PPHHDTPCTCKKTTTCKTCT-----KKTTPVPTPS--SSTESSAPV-----PTPSSS 335  
QY 246 TT--TKGPAITPKE-----PPTTPKEPASTTPKEPTPTTIKSA-----TTPKEPAPT 293  
Db 336 TTSSAPVTSSTSSAPVPTPSSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSSAPV 395  
QY 294 TTKSAPTTPKEPAP-----TTTKBPAP-----TTPKEPAP-----TTPKEP 329  
Db 396 PTPSSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSSAPVTSSTSS 455  
QY 330 APTTKSAPTTPKEPAPTTPKPAPTTPKEPAPTTPKEPTP-----TTPKEPAPTKE 382  
Db 456 APVTPSSSTSSAPV-----SSTESSAPV-----PTPSSSTSSAPVTSSTSS 507  
QY 383 -----PAP--TTPKEPAPTAPKKPAPTTPKEPAPT-----PKPAPAPT-----T 419  
Db 508 SSAPVTPSSSTSSAP--APTSSSTSSAPVTSSTSSAPVTSSTSSAPVTPSSSTSS 566  
QY 420 TKPSPPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTT-----PKPSPPTTPKEPAP 471  
Db 567 TPTVTSSTSSAPVTPSSSTSSAPVTPSSSTSSAPVTPSSSTSSAPVTPSSSTSSAPV 626  
QY 472 ---TTPKEPAPT--TPKKPAPTTPKEPAPTTPKEPAPTTPKPAAPKAPAPTTPKETAP 527  
Db 627 TSTTSSAPVPTPSSSTSSAPVPT-----PSSSTSSAPVTPSSSTSSAPVTPSSSTSSAP 682  
QY 528 TTPKKLTPTTPKLAAPTTPKEPAPTTPKEPAPTTPKEPTP-----TTPKEPAPT--TPKAAAP 583  
Db 683 VT-----SSTESSAPV-----SSTESSAPV-----PTPSSSTSSAPVTPSSST 730  
QY 584 NTPKEPAP-----TTPKEPAPTTPKEPAPTTPKEPAPT--TPKGTAPTTPKAPAPTTPK 637  
Db 731 ESSAPVPTPSSSTSSAPV-----SSTESSAPVTPSSSTSSAPVTPSSSTSSAPVTPSSST 786  
QY 638 APKELAPTTTPKSTPTSTSK-----PAPTTPKGTAPTTPKEPAPT-----TPKEP 682  
Db 787 TESSAPVPT--PSSSTSSAPVPT--TPSSSNITSSAPSTTPSSSTSSAPVPTTPPS 843  
QY 683 APTTPKGTAP-----TTLKEPAPTTPKPAKPAKLAPTTPTKGP-----TSTSDKPAPTTPK 733  
Db 844 SSTTSSAPVSSSTSSAPV-----PTPSSSNITSSAPSSIPSSSTSSSTSGTCT-- 897  
QY 734 ETAPTTPKEPAPTTPKPAAPTTPET--PPPTTSEVSTPTTPKEPTTIHKSPDESPELSA 791  
Db 898 -VTPSSSKYPGSGTETSSTTSSTTTTIVPTKTTTSTVPTTSTTTTTCVSTGNSAGETTS 956  
QY 792 EPTPKALENSKPEPGVPTTKTPAAKPEMTT-----TAKDKTTERDLRTTTPET--TTAA 843  
Db 957 GCSPKVITTT-----VPTTTTSTVSTTTTITVCSGTNSAGETSGCSFKITTTTV 1011  
QY 844 PKMTKETATTBTKTESKITATTTQVTS--TTQDTTPPK-----ITLTKTT 888

```

Db      1012 PCSTSPSETASITPTPTVTVTVTVTSTKPGGEITTTTPTKNIPTTYLT 1071
QY      889 TLA--PKVTT-----TKKILITTEI---MNKPEET---AKPK----- 917
Db      1072 TIAPTFSVTNTFTPTTTTTCSTGNSAGETTSGCSPKTVTTTVPCTGTGTYTTEA 1131
QY      918 -----DRATNS--KATT-----PKPKKPTKAPKK----- 940
Db      1132 TTLVTVATVTITVTSTGNSAGKTITGYTTKSVPTTYVTTLAPSAVTPATNAVPTT 1191
QY      941 -----TSTKPKTM-----PRVRKPKTTPTPRKMTSTMPELN- 972
Db      1192 ITTECSAATNAAGETTSVCSAKTIVSSASAGENTAPSATPTVTTAIPTTVITTESSVGT 1251
QY      973 -----PTSBIAEAM-----LQTTTRNQTPNSKLVENPKSEDA 1006
Db      1252 NSAGETTTGYTTKSIPTTYTTLIFGSGAKNYETVATATNP-----ISIKTTSQA 1303
QY      1007 GGAEGETPHMLLRPHVFMPEVT 1028
Db      1304 TTASASS-----VAPVVTSPLT 1321

RESULT 9
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 09-Jul-2004
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes; pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: UNIPROT:Q041805; EMBL:Z34465; NID:G600117; PID:G600117

Query Match      15.6%; Score 871.4; DB 2; Length 1188;
Best Local Similarity 19.4%; Pred. No. 1.2e-10;
Matches 276; Conservative 90; Mismatches 410; Indels 646; Gaps 49;

QY      10 LLLL--LSVFVIOVSSOE----- 26
Db      17 LLLLAACLACSQVATSEASVIAHRLQAMKEAGGAGOLPADFEEDDRVGAANFN 76
QY      27 -----LSCKGRCFESFERGECDDAQ--CKYDKCCPD----- 58
Db      77 PRLRAYIALQAWHRAFYDPKGYTANWVGEDVCKYNGVICTEALDDPKITVVAGIDLNG 136
QY      59 -----YESFCA-----BEHVSSENO----- 73
Db      137 ADIAGYLPPELGILTLAPFHINTNRFCGIIPKMSRLSLILHDFDVNNRFGVFPYVCL 196
QY      74 -----ESSSSSSS----- 82
Db      197 EMVSLKYLDRNDFEGELPPALFDKDLDAIFVNTNRFGVPIENLGNSTASVIVFANNA 256
QY      83 -----SS 84
Db      257 FVGCIPKSIGRMVKTLDEIIFLNKLDGCLPLEMGLLVNTTVIDVSGNMLVGTIPEQLSN 316
QY      85 SSTIWKIKSKN-----SAANRELQKKLVKDNKKN 115
Db      317 IAKLEQLDVSRNVFTGIVHESICELPALVNFSAFNFFNSEAAVCMFSPDKALVNLDDRN 376
QY      116 -----RTKKK-----PTPKPPVDEAGSGLNDGDFKVTTPDT 147
Db      377 CLGALRPAQKALQCAPVLARPVDCSKHVCAGYETP-----GGG-----PPS 418
QY      148 STTQHNKVTSPKITTAKPINRPSLPNPSDTSKETSITVKNKETTIVETKTTTTNKQTST 207

```

```

Db      419 SPVCPKPAASADM-----PSPHTPP--DVSPPEP----- 444
QY      208 DGKEKITSAKETQSTIEKTSAKDLATSKVLAKPTTAKETTTKGPALTTPKEBPTTP-KE 266
Db      445 -----LPEPGFVPAPAPMPMETPHSPDDADDVVPPTPPVPGKS 481
QY      267 PASTTP-----KEPTPTTIKSAPT-----TPKEPAPTITTKSAPT--TPKEPAPTITTK 311
Db      482 PPATSEFSQVQPPAASTPPPSLVKLSPPQAPVGSPPPVKITSPPAPIGSPSPFPFVSVV 541
QY      312 EPAPTTPKEPAPTITTKKEPAPTITTKSAPTTP-KEPAPTTPKKPAPTTPKEPAPTTPKEPPT 370
Db      542 SPPPPVKSPPPPAPVGSPPPEKSPPPAPVASPPPPVKSPPPPTLVASPPPPVKSPPPP 601
QY      371 TTPKEPAPTTPKEPAPTTP-KEPAPTAPKKPAPTTPKEPAPTTPKEPAPTTPKEPPT 425
Db      602 APVASPPPPVKSPPPPTFVASPPPPAPVASSPPPMKSPPPPTPVSSPPPEKSPPPPPPA 661
QY      426 --TTPKEPAPTITTKSAPTITTKAPAPTITTKSAPTTPKEPSPTTKKEPA--PTTPKEPAPTTP 482
Db      662 KSTPPPEEYPT-----PPTSVKSSPPPEKSLPPPTLIFSPPPQEKPTPSTESKP-PSFP 715
QY      483 KKPAPTTPKEPA---PTTPKE-----PART 504
Db      716 EKESP--PKEPVSPPTPKSPSPPEAPVSSPPPTPVSSPPALAPVSSPPSVKSSPPPA 773
QY      505 TTKKAPTAPKEPAPT-----TPKETAPTTPKKLTPTTPKLAAPTTPKEKAPTTPPEEL 557
Db      774 SSPPFAPQVKSPFPVQVSSPPPAKSPPLAP-----VSSPPQVEKTSPPAPJUSSPPL 828
QY      558 AP-----TTEBPTPTTPEEAPTTPKAAAPTTPKEPAPTTPKEPAPTTPKEPAPTTP 599
Db      829 APKSSPPHHVSSPPPVKSSPPPAVSPPLTPKPASPAHVSPPEVVKSTP--PAP 886
QY      600 TTPKEPAPTTPKETAPTTP-----KGTAPTTLKEPAPTTPKKPAPKELAPTTPKPT 651
Db      887 TTVISP--PSEPKSPPTPVSLPPPIVKSSPPPAVSSPPMTPKSSPP-----PVVSSPP 941
QY      652 STTSKAPAPTTPKGTAPTTPKEPAPTTPKE--PAPTTPKGTAPTTLKEPAPTTPKKAPKE 710
Db      942 PTVKSSP--PPAPVSSPPATPKSPPPAPVNLPPPEVKSSPPPTPVSSPPPAKSSPP-- 998
QY      711 LAPTTTGGTSTTSKAPAPTTPKETAPTTPKAPAPTTPKAPAPTTPETPTTSEVSTPT 770
Db      999 -APMSGPPPEVKS--PPPPAPVSSPPPVKSGPPPPAPVSSPPPVKSPPPAPVSSPP 1055
QY      771 TTKPTPTIHKSPDESTPELSAETPKALENSKPEPGVPTTKTPAATKPEMTTAKDKTTE 830
Db      1056 PVKSPSP--PAPISGPPPPVKSPPPAPVSSPP-----PPVKSGPPPPAP----- 1097
QY      831 RLRLTPTPTTAAPKMTKETATTTEKTTESKITATTQVTSTTTQDTPFKITLKTTL 890
Db      1098 ----- 1097
QY      891 APKVTTTKTITTTTEIMNKPEETAKPKDRATNSKATTPKQPKTKAPKPTSTKPKTMP 950
Db      1098 -----VSSPPP--PIKSPPPAPVSSPPAP 1121
QY      951 RVKPKTTPTRKMTSTMPELNPTSRIDAEAMLQTTTRNQTPNSKLVENPKSEADAG--- 1007
Db      1122 --VKPSSLPPAPVSSPPPVVTFAP-----PKKEQSLPP 1154
QY      1008 GAGETPHMLLRPHVFMPEVTFDMDYLPRVFNQGIINPMLS 1049
Db      1155 PAESOPP-----ESFN-----DILPPIMA 1174

RESULT 10
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25697

```



QY 92 ----- 91  
Db 686 AGSGSGITLPTTILEPKIEGSKKASGVWTEEDGDEDLMEGSGSWSTTINGTGITGS 745  
QY 92 -----KSSKNSAANRELQKLKV-KDNK----- 113  
Db 746 PRSEGTIRVRITTLGDEGPETATKPGISAPDKTGEKSTESDGEKLTVEKDGEAQSS 805  
QY 114 -----KNRIKKPTPKPPVVDEAGSL-----DNGDFK 141  
Db 806 GGSATSGKKEATSGSSSSAKSGTGESEAGSSGSSGSGVSGESVSTESGSGF 865  
QY 142 VTT-----PDSITQHKNVST-----SPKIT 162  
Db 866 TSSSGVSGEATSGTGDSESGKPSKSTEEKLPFTKNGEKSPISGSDTTGKSEBEIT 925  
QY 163 TAKPINRPSLPNS-----DTSKETSILTVNK----- 189  
Db 926 SRKPIRGSDSLTEGSGSGEWFTETGSKGHFESGSKSVTSGKPTQSGAEGSGGPKVPKG 985  
QY 190 ----ETTVETKETTT-----NKQSTDGKEKTTSAKETQ 220  
Db 986 PGAPELITDGESSSTSGDKSGGKADKSDNNKVPKTDGKNPDIITDGEDSTSETSGGE 1045  
QY 221 SIEKTSADLAP-----TSKVLAKPT----- 241  
Db 1045 QGPKGSKGQPPGDGSEVKKPTSEVDGPNLSGKGNVPLKPTDLPBEGSGILITSS 1105  
QY 242 ----PK-----ABTTKGPAITPKBPTPT--TPKE----- 266  
Db 1106 GGNKSTFEHGTKLERLPFKTEDKSSBTPQLGLEISAGKPEPDGTSKEVGLIEWESTT 1165  
QY 267 PASTT----- 271  
Db 1166 POSTILDSVGLIEISGDLTKATKPHVEIEGSGTGDEIEITATTRDVSXSKTKKPRVEVD 1225  
QY 272 ----PKEPTTIKSA-----PTT-----PKEP----- 290  
Db 1226 GGDNGETSGVGGKPTTAPTPSSAESSTRIPTTSEASPEGSGEAGVPESPDGSGESS 1285  
QY 291 ----ATTTKSAPTTKEPAPT-----TKEPAPTTPKEPAPT----- 325  
Db 1286 TSAPOGVSTSSATAPEVPTTSASSTPDAAVEESGIPSTKPTAEPLETT--APSTEV 1341  
QY 326 ----TKEPAPTTKSAPTTPKEPAPT-----KKAPAPTTPKEPAPT----- 364  
Db 1342 SPFGSGTEESTLPPTEGSGESTTSSAPT--EPATVLPQNRNEKPEPTKDTFALPTTT 1399  
QY 365 ----PKE----- 367  
Db 1400 GAPQANDSSVENTKCTSSDEGLDALCERRTGVRCRCEPGFEGAPPKSCVDVDECATGDH 1459  
QY 368 ----PT----- 369  
Db 1460 NCHESARCONVGGYACFCPTGFRKADGSCQDIDECTEHNSTCCGANAKVKNKPGTYS 1519  
QY 370 ----PTTPKEPAPTTK----- 381  
Db 1520 ECENGFLGDGYQCVPTT--KKPCDSTQSSKSHCSNMSCEVDTVDSGVECKEOMGGYKKS 1578  
QY 382 ----EPAP----- 385  
Db 1579 GKVCEDINECVAEKAFCSLNANCVNMNGTFSCSKQGYRGDGPMTCTDINECDERHPCPH 1638  
QY 386 ----TPKE----- 390  
Db 1639 AECTNLEGSFKCEHSGFEGDGKIKCTNPLERSCEDVEKFCGRVDHVSCLSVRLYNGSL 1698  
QY 391 ----PA----- 392  
Db 1699 SVCEBPGFRFEKESNCVDIDECESRNNCDPASAVCVNTEGSKYRCECABEGVEGGVC 1758  
QY 393 ----PTAPK----- 397

Db 1759 TDIDCDRCMAGCDSDMAMCINRMGSCGCKCMAGYTGCGATCIKIBEEPKSDKTADEWS 1818  
QY 398 ----KPAPTTPKEPAP----- 409  
Db 1819 RLCELEKKQCTVDEEVRVPOCGACLPGHHPINGTCQSLQISGLCAQKNDCKNHAECIDIHP 1878  
QY 410 ----- 409  
Db 1879 DSHFCSCPDGFIDGMI CDDVDECNAGMCDDEKCENTIGSFNCVLCLEGFKKVDKCV 1938  
QY 410 -TTPKEP-----APTTPKSPPTTPKEPAPTTPKSAPTTPKAPAPTTPKSAPT 456  
Db 1939 VDEKKQPNREKIEIDRENSSSNSGQEKPTTKGIVSSTSATSESTIAPHVITISSTT 1998  
QY 457 TPKEPSPTTPKEPAPTTPKEPAPTTPKAPAPTTPKAPAPTTPKAPAPTTPKAPAPTAPKE 516  
Db 1999 STKDMTSSKSPENVMTMSSEPEVSTSSSKSTTASETTIVSSTPSESS-----SSEAPLT 2051  
QY 517 PAPTTPKETAPTTPKLTPTTPE-----KLAPTTPE-----KPAPTTPPEELAPT 561  
Db 2052 SSPAITTEV--ITESVSKSTTPKESSESEITVKLSSKSPTEVSSVKSPSTPS-----TT 2105  
QY 562 PBEPTPTTPEEPAPTTPKAAAPNTPKEP-----APTTPKEPAPTTPKAPAPTTPKET-- 613  
Db 2106 SQSVTSTVPETSKSTVLSSEAPVTSTSPTEVHTSETKLSLSASSTGDNSTTPSTSSL 2165  
QY 614 ----APTTPKGT-----APTTLKEPAPTTPKKAPKAPELAPT-----TTK 648  
Db 2166 ASVKSTSAPETSASVAPVKLSLSPDV--SQPSTKTFDATESTVQASSETSSOTSVKSTS 2224  
QY 649 EP-----TSTTSDKAPAPTTPKGTAPTTPKEPAPTTPKAPAPTTPKG----- 689  
Db 2225 EPESHVKLSITSSNPSVPTSPKSTPTPEST--EQPTSTTPSGQSLTPMNSNEVL 2282  
QY 690 ----TAPTTTLKEPAPT--TPKKAPAKEL--APTTLTKGPT----- 720  
Db 2283 TTSEPHVLSLSPDVQSSTTPNNLSESTVETPKTSSEVLSNSEPSTTEAPTTLSPD 2342  
QY 721 --STT-----SDKP-----APTTP----- 732  
Db 2343 ILSTTTNNLSQSSTVSTEDRSEISSENSEKPTSAPELVTSSVTHVASSPDVPTESSEP 2402  
QY 733 ----KETAPTT-----KEPAPTTPKP-----APTTP-- 757  
Db 2403 DLTGSTENIPEASSKQSTISPTPTDTTTTASEPTKSTSMSPDLSTSNVLSSESSTPES 2462  
QY 758 ----TPPPTTSEVSTPTTKBPTT 777  
Db 2463 SSKSPVSSSTEGISVVVSTSEFSKVPESTISSVLEEDLTKTTPSPILEETTTASETSEPLT 2522  
QY 778 ----IHK--SPDESTPELSABPTPKALENSP--KEPG-----VPTT----- 810  
Db 2523 EDSLATVSVRIHELTTSENVPKSESTTTSSSESKPSQEPAGILTSTVVVPTSSVSLITA 2582  
QY 811 ----KTPRAATKPEM-----TTAKDKTTERDLRTTPTETTAAPKMT 847  
Db 2583 SEIEAITSNTPPKQGRPTITTPSKSLVKSTTSPSTVTSSEPESEKRTTSTVSTVTTPT 2642  
QY 848 KETATT-----TEKTTESKITATTITQVTS-----TTTQD 876  
Db 2643 EBITTSESILLTAAPSKPTTESTTESEAPTTTAKTSEKPSNVVSTSRKSTENVEITSQ 2702  
QY 877 TTPFKITTLKTT-----TLAPKVTTTKKTTITT----- 904  
Db 2703 SGLSLESSTMSSTSRBPETNAPAVTVSSSEASSTTLEENSTSSPTSSEASVKLSLPPESI 2762  
QY 905 -----EIMNKKPEE----- 912  
Db 2763 TSEAVTVSSRAPAEITMSESHREISTVSSPESEPEIPLSTTVSPNVVITASSIPSEPI 2822  
QY 913 ----TAKPKD-----RATNSKA-----TTP 928

```

Db      2823 SSVTSSTPRVRLITCTPDDLIIVSVTPSHGNRRQNTASSVPSNSTSPILLPSESILTP 2882
Qy      929 KQOKPTKAKPKTSTTKKPTMPVRKPK-----TTPTP-----RKMTSTMP 969
Db      2883 QPPTTTTAKPATTSGKGPSPQIOPFAEMFTTTPAPPSPNGGYGGEETNOEBQVIST-- 2940
Qy      970 ELNPTSRIFAAMIQT-----TTPRPNCTPNSKLVENVNPKS 1003
Db      2941 -----TTIEPSLCSTVTCHSLATCEOSTGVCICRDPFIGDGTACSKKSTADCSISLSC 2996
Qy      1004 EDAGGAEGET-----PH---MLLRPHVPMPE 1026
Db      2997 ADKAKCDNSRSCEDAGYIGDGVCSHPQDCVLRDNLCSPE 3039

RESULT 12
T29757
protein UNC-89 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C;Accession: T29757
R;Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid C09D1.
A;Reference number: Z20679
A;Accession: T29757
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-6642 <DUZ>
A;Cross-references: EMBL:AF003131; PIDN:AA854132.1; GSPDB:GN00019; CESP:unc-89
A;Experimental source: strain Bristol N2; clone C09D1
C;Genetics:
A;Gene: CESP:unc-89
A;Map position: 1
A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match      15.4%; Score 857.5; DB 2; Length 6642;
Best Local Similarity 16.5%; Pred. No. 2.4e-09;
Matches 366; Conservative 159; Mismatches 399; Indels 1295; Gaps 83;

Qy      9 YLLLLLVFVIQVSSQE-----LSCKGRCFESPERGEC---DCDAQCKK 51
Db      393 YSSIRLDKYNIRQHTTDEDIVLQPOEPGLPSRIKPKDFETSEYVRKAWLRDIAEBQEK 452
Qy      52 YDKCCPDYEFCAEHSV-----ENQESSSSSSSSSTIIV----- 89
Db      453 Y-----AAERDAISMATSEMTASSVDPMNASDQQSFSEWSGSRKSLPFGPE 502
Qy      90 -----KIKS-----SKNSAANRELQK 105
Db      503 EGGPPRKVKSPVVISPTGSSTSIYSGGSSIDWTTTGTLEMQGTRVTRQYGFRTLOE 562
Qy      106 ----- 105
Db      563 SSAKMCLKVTGYPLDITWYKDDVQLHEDRHTFYSDEGFFAMTIDPVQVDTGRYTCM 622
Qy      106 ----- 105
Db      623 ATNEYQASTSAFFRVLKVEKAAPFAVTKLRDKECKEGBVDIDFECEVGEWPEPELVL 682
Qy      106 -----KLKVKDNK----- 113
Db      683 VDDQPLRPSHDFRLQDGGTAKLEIRDAQPDGTVTVTKIQNEFGSIESKAELFVQADPD 742
Qy      114 KN----- 115
Db      743 KHVAPBFQATIEYVECDGEVEFRFKSVITGDPNPELIWFINGKPLSESKVKFISBDGI 802
Qy      116 -----RTKKKPTPKPPVVD----- 130
Db      803 CIITIKDVTFRHFGMVTTCQSNRLGSCDGRUKVRVPPAPPTFNKPLEDKTVQEKSTVV 862

```

```

Qy      131 -----AGSLDNGD-----FKVTTPTDSTTOHN----- 153
Db      863 FEVDVSGNPEPTLFTLCKGKXGEGVEIVGHDGFYRISIFNTSMXKHDGELIVAKAQN 922
Qy      154 -----KVSTSPKITTAKPI-----NPRPSL----- 173
Db      923 EHGTAESRARLTVQEQUEEERSAP--TFLKDIEDQTVKGFVAVFETTVRGNENPEVTFW 980
Qy      174 ----- 173
Db      981 INGHKMDQSGPGVKIEAHNHDKLITIDSAQYACTVLCRAENAVGRFETKARLVVLAPEKQ 1040
Qy      174 --PPNSDTSKETSLTVNKETT-----ETKETTT----- 200
Db      1041 KKPP-----KFVEILLVDKTTETVDNTVVFVRVEGEPKPTVWYLKGEELKQSDRVEIREF 1095
Qy      201 -----TNKOTSTDGKEKTTSAKETOSIEKTSIAKDLAPTSKVL 237
Db      1096 DGSIKISIKNIKIEDAGEIRAVATNSEGSDETAKLTVQKKPAPE---FDLRPVSLTV 1151
Qy      238 AK-----PTP-----KAETTTKGPALTT----- 255
Db      1152 EKGSEAVFSAHAFGIPLPTYWMSVNGRKVRDQOGGARVTRDESTVDGASILTIDTATYYS 1211
Qy      256 -----PKEPTTTPKEPAST-----TPKE 274
Db      1212 EVNHLTISVVAENTLGAETGAQLTIEPKKESVVVEKQDLSSESVQKEIAQQVKEASPEA 1271
Qy      275 PTTTITKSAPTTPKEPAPTT-----KSAPT----- 300
Db      1272 TTTTITMETSILTSKTTTMSITTEVTSTVGTVETKESESATVIGGGSGGVTEGSISV 1331
Qy      301 -----TPKEPAPTTTKEPAP-----TTPKE-----PAPTTTKEPAPTTTK 335
Db      1332 SKTEVVSKTDSQDQVDEGTPKRRVSAFEBELPKVEIDSRKKKSSP--DKKESSEKTE 1390
Qy      336 SAPTTTKEPAPTTPKKAPATTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 393
Db      1391 EKPASPTKKTGEEVKSP---KEKSPASPTKKEKSPAAEEVKSPTKKEKSPSSPTKKEKSP 1447
Qy      394 TAPKAPATTTPKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 434
Db      1448 SSPTKKTGDEVKEKSP--PKSP---TKKESPEKPEDVKSPVKKEKSPDATNIVEVSSET 1502
Qy      435 TTKSAPT-----TKEPAPTTTKSAPT-----KEPSPTTTKEPAPTTT 474
Db      1503 TIEKTETMTMTHESESESRTSVKKEKTPKVEDEKPKTKKDKKSPKSIIEIKSPVK 1562
Qy      475 KEPAP--TTPKKPAPTTTPKEPAPTTTPKEPAPTTTTPK-----PAPT-----A 513
Db      1563 KEKSPKEVEKSPASPTKKEKSPKSPASP---TKKSENEVKSPTKKEKSPKSVVEELKS 1618
Qy      514 KPAPATTTPKEPAPTTPKKLTPTTPEKLAPTTPEKPA-----PTTPEELAPTTPEEPTP 567
Db      1619 PKEKSP---EKADDPK--SPTKKEK---SPEKATEDVKSPTKKEKSPKVEEKPTS 1668
Qy      568 TTPKEEPA--TPKAAAPNTPEKAPTTTPEK--PAPTTTKEPAP-----TTPKEAPT 616
Db      1669 PTKKESPTKKTUDEVKSPTKKEKSPQTVEEKPASPTKKEKSPKSVVEEVKSPKESKSP-- 1727
Qy      617 TPKGTAPTTLKEPAPTTPKKAPKELAPTT-----KEPTSTTSDKAPPTTPX 664
Db      1728 -----EKAEKSPSPTKKEKSPKSAEEVKSPTKKEKSPKSAEAEKSPKPTKK 1776
Qy      665 GTAPT-----TPKEPAPTTTPKEPAPTTTPKGTAPTTLKEPAPTTPKKAPKELAPT 714
Db      1777 ESSFVRMAADDEVKSPTKKEKSPKEVEE--KPASP---TKKE---KTEKSAEELKSP 1826
Qy      715 TTK-----GPTSTTSD-----KPAPTTTPEKAPTTTTPKEPAPTTPKKAPPTTPE 757
Db      1827 TKKEKSPSPTKKTGDESKESPEKPEEKSPKSPTKKSPGSPKPKKSKSPKAEKPAKP 1886
Qy      758 TTPPTTSEVSTPT-----TTKEPTTIHKSPD----- 783

```



Db 1887 ----LRFDLKLOVWNTDLAHPFVWVEHATECKWFLDGKELITTAOGVTVSKDQDFRCS 1942  
QY 784 ----ESTPELSAEPPTKALENSPKPGVPTTKTPAATKPEMT 821  
Db 1943 IDTFMGSGTVSVASNAAGSVETKLETPK-----ETK-----KPEFT 1985  
QY 822 TTAKD-----KTERD----- 832  
Db 1986 DKLRDMEVTKGDTVQMDVIALHSPLYKWYQGNLLEDGKNGVYIKNEENKSSLIIPNAQD 2045  
QY 833 ----LRTTPETTTA-----APKMT-----KETATTTKTTES----- 860  
Db 2046 SGKITVEASNEVSGSSSAQLTVNPPSTTPIVVDGPKSVTIKETAEFAFKATISGFPAPT 2105  
QY 861 ----KITATTQVTSITTTQDITPKIT-----TLKTT----- 888  
Db 2106 VKWTINEKIVEESRTITTKTEDVYTLKISNAKIEQTGVKVTAAONSAGDSKQADLKVE 2165  
QY 889 --TLAPKVTTTKTIT----- 2165  
Db 2166 PNVKAPKF-----KSLQTDKVADEGEPLRWNLLEDGSPGTEVSWLLNGQPLTKSDTVQVVD 2222  
QY 909 ----KPEE-----TAKPDRA-----TNSKAITPKQKPTKAPKETS----- 942  
Db 2223 HGDGTYHVTLAEAKPEMSGTLTAKAKNAAGECETSAAKVTVNGGKPKPFVQAPQNHETTL 2282  
QY 943 ----TKYP-----KTMFRVRKP----- 955  
Db 2283 EESVKFSIAVITGKPMNVWYLNKKLIQSEEVKVYVHETGKTSIRIQKPLMEHNGTIR 2342  
QY 956 ----KTTPTRKWTSTM-----PE----- 970  
Db 2343 VEAEVSGKVQATAQLKVDKKTVP-KFTYNDMDRQKGEDEVKFTANVEGYEPVAVWT 2401  
QY 971 LN--PTSRIAEMLOTTTPRNPQNSK-----LVEVNPKSEDAGG----- 1008  
Db 2402 LNPVPSK-----HPNITVTDKGBHTTEISAATPEQAGELSCAETNPVSGSKRD 2451  
QY 1009 ----AEGETPHMLLRPHVFP-----EVTPEMDY 1033  
Db 2452 VQIAVKVKGDAPTFAKNLEDRLITEGELTMDAKLINVKPKRITWLKDGVEITSDGHY 2510

RESULT 13  
B48666  
cell proliferation antigen Ki-67, short form - human  
C:Species: Homo sapiens (man)  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 15-Mar-2004  
C:Accession: B48666  
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde  
J. Cell Biol. 123, 513-522, 1993  
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiqui-  
ous.  
A:Reference number: A48666; PMID:94043435; PMID:8227122  
A:Accession: B48666  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-2897 <SCH>  
A:Cross-references: EMBL:X65551  
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat  
F;29-91/Domain: Kinase interaction domain homology <KIH>

Query Match 15.0%; Score 835.4; DB 2; Length 2897;  
Best Local Similarity 15.3%; Pred. No. 2.1e-09;  
Matches 392; Conservative 138; Mismatches 420; Indels 1521; Gaps 94;  
QY 22 VQSQELSCK-----GRCFESFERGECDCD 46  
Db 382 ISNQKDFEDLSGIAEMFPTVKEQPQLTSTCHIAISNENLLGQFQGTDSGE----- 437  
QY 47 AQCKYDKCQPDYEST-----CAEEH-SVSE----- 71

Db 438 -----PLLPSTSESGNVFFSAQNAAKOPSDKCSAPPLRRQCIRENGVNAKTPNTY 490  
QY 72 ----NQSSSSSSSSSSSTIWKIKSSKNSAANBELQ----- 105  
Db 491 KMTSLETKSDTETEPSTVSTVNRSGRSTEFNRNIQKLPVESKSEETNTEIVECILKRGQ 550  
QY 106 ----KLKVKON-----KKNET-KKKPTPKPPVUDEAG 132  
Db 551 KATLLQORREGEMKEIERPETTYKENIELKENDEKWKAMKRSRTWGQKCAPMSDLTD----- 607  
QY 133 SGLDNGDFKVTTPDT-----STTOHNVSTSPKITTAKPINPRPLPP-----N 176  
Db 608 ----LKSLEPDTLMKDTARGQNLQTDHAKAPKSEKGIK-----MPCQSILQPEPIN 657  
QY 177 SDTS-----KETSILTVMKETTIVETKETTNTNNKOTSTGKE----- 211  
Db 658 TPTHKQQLKASLGKVGVEELLAVGKFTFT-SGETHTHREHPAGDCKSIRTFKESPKQI 716  
QY 212 ----KTTSAKETQSIEKTSAKDLAPTSKVLAKPTKAE----- 245  
Db 717 LDPAAVTCMKKWPRTPKEEAQSLE-----DLAGFKELFOTPGPSEESMTDEKTTKIACK 771  
QY 246 ----TTTK-----GPALTTPKEP----- 259  
Db 772 SPPPEVDPTSTKQWPKSLRKADVEEFALRLKLTSPSAGKAMLTTPKAGGDEKIDKAF 831  
QY 260 --TPT-----TPKEPAST-----TPKEP 275  
Db 832 MGTVPQKLDLAGTLPGRKQLQTPKEKAQALDLAGFKELFOTPGHTEELVAAKTTKIP 891  
QY 276 ----TPTTIKSAPT-----TPKEP----- 292  
Db 892 CUSQSDPDVDTPTSTKQPKRSIRKADVEGELLACRNLMPSAGKAMHTPK-----PSVGECK 948  
QY 293 ----TTTKSAPTTPKEPAPT----- 308  
Db 949 DIIIFVGTVPQKLDLTENLTGSKRRPQTPEKAAQALDLTGFKELFOTPGHTEEVAAGK 1008  
QY 309 TTKEPAPTTPKEPAPTTPKEPAPTTPKESAPTTPKE----- 343  
Db 1009 TTKMPCSSPPESADT-----PTSTRQPKTPLEKRDVQKLSALKLKTQTSGETHTD 1062  
QY 344 ----PAPTTPKAPPTTPKEP----- 361  
Db 1063 KVPGEEDKINAFRETAKQKLDPAASVTSKSRHPT-KKAAQPLEDLAGWKELFOTPVCT 1121  
QY 362 --PTTPK-----EPTPT-TPKEPAPTTPKE----- 382  
Db 1122 DKPTTHEKTTKIACRSQDPDVTPTSSKPSQSKSLRKVDVEEFFALRKRTPSAGKAMHT 1181  
QY 383 PAPT-----TPKEP----- 392  
Db 1182 PKPAVSGEKNIYAFMGTVPQKLDLTENLTGSKRRRLQTPKEKAQALDLAGFKELFOTRGH 1241  
QY 393 ----PTAPKPEP----- 403  
Db 1242 TBESMTNDKTAQVACKSSQPDLDKKNPASSKRRRLKTSLGKVGKVEELLAVGKLTQTSGETT 1301  
QY 404 PKEPAPT-----TPKEP-----P 417  
Db 1302 HTHTEPTGDKGKMAFMESPKQILDASAASLTGSKRQLRTPKGSKEVPEDLAGFIELFOTP 1361  
QY 418 TTTKPESPPTPK-----EP-----APTTPKSAP----- 440  
Db 1362 SHTK-SMTNEKTTKVSRAASQPDVDTPTSSKQPKPKRSRKADTEEEFLAFKOTPSAG 1420  
QY 441 ----TTTKEP----- 447  
Db 1421 KAMHTPKPAVGEKINDINFLGTPVQKLDQPGNLPQSNRRLOTRKKAQALELTGPRELF 1480  
QY 448 ----PT-----SAPTTPKPSPTTKE-----PA 470  
Db 1481 QTPCTDNPTADEKTTKILCKSPQSDPADTPTNTQPKRSLKADVEEFLAFKLTFS 1540

QY 471 -----PPTPKPAPT----- 480  
Db 1541 AGKAMHTPKAAVGEEXDINTFVTPVEKLDLLGNLPGSKRRPQTPKEKAKALEDLAGFKE 1600  
QY 481 ---TP---KKPAPTTPKEPA----- 494  
Db 1601 LFTQPGHTEBMTDDKI TEVCSKSPQDPVKTFTSSKQRLKISLGKVGUVKEEVLVPGKLT 1660  
QY 495 -----PPTPKPAPA----- 502  
Db 1661 QTSKGTQTHRETAGDGKSIKAFKESAKQMLDPANYGTGMRWPTPKBEAQSLDLAGF 1720  
QY 503 -----PTTKKAPAPTAPKEPAP---TTPKET--APTTP-----KK 532  
Db 1721 KELFQTPDHTBESTDDKTTKIAKSPPPESMDTPTSTRRRPKTPLGKRDIVBELSALKQ 1780  
QY 533 LTPPTT-----PEKLAP---TTPKEKPAPTTP-----BELA--- 558  
Db 1781 LTQTHWDKVPGEDKGINVERETAKQKLDPAASVTGSKBQPTTPKGAKQPLEDLAKLE 1840  
QY 559 -----PTTPEBPTTPPEPAPT----- 576  
Db 1841 LFTQPVCTDKPTTHTXHTTKIACRSQPDPVGTPTIFKPQSKRSLRKADVEESIALRKRT 1900  
QY 577 -----TPKAAA-----PNTPKPEPA----- 590  
Db 1901 PSVGKAMDTPKPAAGGDEKDMKAFMGTFVQKLDLGNLPGSKRWPQTPKEKAQALEDLAGF 1960  
QY 591 -----PTTPKEPAPTTPKEPAP---TTPKET----- 613  
Db 1961 KELFQTPGTDKPTTDEKTKIACKSPQDPDVPDTPASTKQRPKNLKNKADVEEBFLALRK 2020  
QY 614 -----APTTPK-----GTAPTTLKEPAPTTPPKAP----- 639  
Db 2021 TPSAGKAMDTPKPAVSDEKNINTFVETPVQKLDLGNLPGSKRQ--PQTPKEKALEDL 2078  
QY 640 ---KELAPTTKEPTSTTSK-----PAP----- 660  
Db 2079 VGFKELFQTPGHTESMTDDKITEVCSKSPQSPESFKTSRSKORLKIPLVKVDMKSEPLA 2138  
QY 661 -----TTPKGAPT-----TPK-----EPAPT-----TPKEPA--- 683  
Db 2139 VSKLTRTSGETTQHTTEPTGDSKSIKAFKESPKQILDPAASVTGSRRLKTRKEKARALE 2198  
QY 684 -----PTTPKGAPTTLKEPAPTTPPKAPAPKEL 711  
Db 2199 DLVDFKELFSAPGHTBESMTIDKNKIPCKSPPPPELTDATSTKRCF-KTRPRKEVKEEL 2257  
QY 712 AP---TTTKGPTSTTSKPA-----PTTPKE----- 734  
Db 2258 SAVERLUTQSGQSTHTHKEPASGDGHIKVLQRAKKKPNVVEEPSRRRPRAPKEKAQPL 2317  
QY 735 -----TAPTTPKEPAPTTPPKPAPTTPPETPTTSEVSTPTTKE 774  
Db 2318 EDLAGFTELSETSGHTQESLTAGKATKI PCESPPLEVVDITASTKRLHTRVQKVQKEE 2377  
QY 775 PTHIHSPEDESTELSAEPTP-----KALENSPKPEPGVPTT----- 810  
Db 2378 PSAV-KFTQTSGETTDADKEPAGEDKGIKALKESAKQTPAPAASVTGSRRRRPRAPRESAQ 2436  
QY 811 -----KTPAA--TKPEMTTAKTKTERDLRTPE---TTTAAPKMTKETATTEKT 857  
Db 2437 AIEDLAGFKDPAAGHTEESMT---DDKTTKIPCKSSPELEDATSSKRRPRTPAKQVEVK 2493  
QY 858 TE-----SKITATTTQVSTTQ-----DTTPPKITTLKTTLAPKVTTKKTIITTEIM 907  
Db 2494 EELLAVGKLTQTSGETTHDKPEVGEKGTKAFK-----QPAKRNVDABDV 2540  
QY 908 NKPEETAKPKDRA-----TNSKATTPKPKQKPTK-- 935  
Db 2541 GSRQRPRAPKEKAQPLEDLASFQELSQTGHTTEELANGAADSFTSAPKQTPDSGKPLKIS 2600

QY 936 ---APK-KPT----- 941  
Db 2601 RRVLRAPKVEPVGDVVSTTRDPVKSQKSNSTSLPPLPFKRGCGKSGVGTGKRLRCMPAPE 2660  
QY 942 -----STKKPTMRVRKPKTPTT---PRKMTSTMPBLNPTSRIAEAMLQTTTPNQ- 990  
Db 2661 EIVEELPASKQKQVAPRARGKSSEPVVIMKRSRLTSAKRIEPAEELNSDMKTNKEEHL 2720  
QY 991 ---TPNSKLVEVNPXKSDAGAGETPHMLL-----RPHFMEVTPDMDYLP 1035  
Db 2721 QDSVPENKGISLRSRQDKTEAQOQITEVFLAERIEINRNEKKP---MKTSPEMD--- 2773  
QY 1036 RVENQGIINP 1046  
Db 2774 ---IONP 2777  
RESULT 14  
A48666  
cell proliferation antigen Ki-67, long form - human  
C:Species: Homo sapiens (man)  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 09-Jul-2004  
C:Accession: A48666  
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdes  
J. Cell Biol. 123, 513-522, 1993  
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq  
ins.  
A:Reference number: A48666; MUID:94043435; PMID:8227122  
A:Accession: A48666  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-3256 <SCH>  
A:Cross-references: UNIPROT:P46013; EMBL:X65550; MID:g415818; PIDN:CAA46519.1; PID:g4158  
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat  
F:29-91/Domain: kinase interaction domain homology <KH>  
Query Match 15.0%; Score 834.9; DB 2; Length 3256;  
Best Local Similarity 15.9%; Pred. No. 2.5e-09;  
Matches 392; Conservative 138; Mismatches 420; Indels 1521; Gaps 94;  
QY 22 VSSQELSK-----GRCFSEFERGECDCD 46  
Db 741 ISNQKDFDLGSLGIAEMFKTPVKEQPLTSTCHIAISNENLLGKQFGTDSGEE--- 796  
QY 47 AQCKKYDKCCPDYEF-----CAEEH-SVSE----- 71  
Db 797 -----PLLTSESFGNVFFSAQNAAKQPSKASPLRRQCIRENGNVAKTPNTY 849  
QY 72 ---NORSSSSSSSSSTIWKIKSSKNGSAANRELQK----- 105  
Db 850 KMTSLETKTSDTETESKIVSTVNRSGRSTEFNIOQLPVESKSEETNTEIVCILKRGQ 909  
QY 106 -----KLKVKON-----KKNT-KKKTPKPPVDEAG 132  
Db 910 KATLQORREGEKEIERPPETYKENIELKENDEKMKAMKRSRTWGCKAPMSDLTD--- 966  
QY 133 SGLDNGDFKVTTPDT-----STTQNKVSTSPKITAKPINRPSLPP--N 176  
Db 967 -----LKSPLDTEMLKMDTARGONLLQTDRAKAPKSEKGIKTK--MFCQSLQPEPIN 1016  
QY 177 SDTS-----KETSITVKNKETTETKETTNNKQSTDKGE----- 211  
Db 1017 TPTHTTKQLKASLGKVGKVELLAVGKFTRT-SGETHTHREPADGKGSIRTFKESPKQI 1075  
QY 212 -----KTTSAKETQSIKTSADLAPTSKVLAKPTKAE----- 245  
Db 1076 LDPAARVTGKMWPRTPKBEAQSL-----DLAGFKELFQTPGPESESMDEKTKTIACK 1130  
QY 246 -----TTTK-----GPALTTPKEP----- 259  
Db 1131 SPPESVDPTSTKQWPKRSLRKADVEEBFLALRKLTSPSAGKAMLTFPKAGGDEKQIKAF 1190  
QY 260 --TPT-----TPKBPAST-----TPKEP 275

```
Db 1191 MGTVPVQKLDLAGTLPQSKRQLQTPKEKAQALDLAGFKELFQTPGHTHEELVAAGKTKIP 1250
QY 276 -----TPTTIKSAPT-----TPKEPAP-----292
Db 1251 CDSQSPDPVDTPTSTKQPKRSIRKADVEGELLACRNLMPSAGKAMHTPK-----PSVGEK 1307
QY 293 -----TTTKSAPTTTKEPAPT-----308
Db 1308 DIIIFVGTVPQKLDLTENLTGSKRRRQTPKEBAQALDLTGFKELFQTPGHTTEAVAAGK 1367
QY 309 TTKEPAPTTPKEPAPTTPKEPAPTTPKKSAPTTPKE-----343
Db 1368 TTKMPCESSPPESADT-----PTSTRQPKTFLEXRDVQKELSALKKLTQTSGETHTD 1421
QY 344 -----PATTPKPAPTTPKEPA-----361
Db 1422 KVPGGEDKINAFRETAOKLDPAASVTGSKRHPT-KEKAQPLEDLAGMKELFQTPVCT 1480
QY 362 --PTTPK-----EPTPT-TPKEPAPTTPKE-----382
Db 1481 DXPTTHEKTKIACRSQDPVDTPTSSKQSKSLRKVDVEEFFALRKTPSAGKAMHT 1540
QY 383 PAPT-----TPKEPA-----392
Db 1541 PKPAVSGEKNIYAFMGCTVPQKLDLTENLTGSKRRRLQTPKEKAQALDLAGFKELFQTRGH 1600
QY 393 -----PTAPKKA-----PTT 403
Db 1601 TEESMTNDKIAKVACKSQPDLDKNPASKRRRLKTSGLGVGKVELLAVGLTQTSGET 1660
QY 404 PKEPAPT-----TPKEPA-----P 417
Db 1661 HHTEFTGDKSMKAFMESPQKILDSASLTGSKRQLRTPKGSFVDPDLAGFIELFQTP 1720
QY 418 TTKEPSPTPK-----EP-----APTTPKSAP-----440
Db 1721 SHTKE-SMTNEKTTKVSYRASQDVLDTPTSSKQPKRSRLKADTEEBFLAPRKQTPSAG 1779
QY 441 -----TTTKEPA-----447
Db 1780 KAMHTPKPVGEEKDINTFLGTPVQKLDQFQNLPGSNRRRLQTRKEKAQALELTGPRELF 1839
QY 448 -----PT-----SAPTTPKESPTTPKE-----PA 470
Db 1840 QTPCTDNPTADEKTKIILCKSPQSDPAPTPTNKQPKRSRLKADVEEFFFLAPRKLTFS 1899
QY 471 -----PTTPKEPAPT-----480
Db 1900 AGKAMHTPKAAVGEEDKINTFVGTVPVEKLDLGNLPGSKRRRQTPKEKAQALDLAGFKE 1959
QY 481 --TP-----KKPAPTTPKEPA-----494
Db 1960 LFGTPGHTBESMTDDKITEVSKSPQDPVKTPTSSKQRLKISLGVGVKEEVLVPVGLT 2019
QY 495 -----PTTPKEPA-----502
Db 2020 QTSKGTQTHRETAGDKSIKAFKESAKQMLDPANYGTGMRWPRTPKEAQSLDLAGF 2079
QY 503 -----PTTPKPAPTAPKEPAP-----TTPKET-----APTTP-----KK 532
Db 2080 KELFQTPDHTESMTDDKTKIACKSPFPESMDTPTSTRRRPKTPLGRKDIVEELSALKQ 2139
QY 533 LPTPT-----PEKLAP-----TTPKEPAPTTP-----BELA-----558
Db 2140 LQQTHTDKVPGDEKGINVFRETAQKLDPAASVTGSKRQRTPKGAQPLEDLAGLKE 2199
QY 559 -----PTTPEEPTTPTEEPAPT-----576
Db 2200 LFGTPVCTDKPTTHEKTKIACKSPQDPVGTPTIFKQSKSLRKADVEEBSLARKKT 2259
QY 577 -----TPKAAA-----PNTTPKEPA-----590
```

```
Db 2260 PSVGKAMDTPKPAGGDEKDMKAFMGTPVQKLDLGNLPGSKRWPQTPKEKAQALEDLAGF 2319
QY 591 -----PTTPKEPAPTTPKEPAP-----TTPKET-----613
Db 2320 KELFQTPGDKPTTDEKTKIACKSPQDPVDTPTASTKQPKENLRKADVEEFFALRK 2379
QY 614 --APTTPK-----CTAPTTLKEPAPTTPKKEPAP-----639
Db 2380 TPSAGKAMDTPKPAVSDEKNINTFVETPVQKLDLGNLPGSKRQ--PQTPKEKALEDL 2437
QY 640 -----KELAPTTTKEPTSTTSDK-----PAP-----660
Db 2438 VGFKELFQTPGHTESMTDDKITEVSKSPQESFKTSRSSKQRLKIPLVKVDMKKEPLA 2497
QY 661 --TTPKGTAPT-----TPK-----EPAPT-----TPKEPA-----683
Db 2498 VSKLRTSGETTQTHTEPTGDSKSIKAFKESPKQILDPAASVTGSSRRQLRTRKEKARALE 2557
QY 684 -----PTTPKGTAPTTLKEPAPTTPKPKAPKEL 711
Db 2558 DLVDFKELFSAPGHTESMTIDKNTKIPCKSPPELDTATSTKRCP-KTRPRKEVKEEL 2616
QY 712 AP-----TTTKGPTSTTSDKA-----PTTPKE-----734
Db 2617 SAVERLTQTSQGQSTHTHKEPASGDEGIKVLQKRAKKPNPVEEPPSRRRPRAPKEKAQPL 2676
QY 735 -----TAPTTPKEPAPTTPKPAPTTPPETPTPTTSEVSTPTTPKE 774
Db 2677 EDLAGFTELSETSGHTQESLTAGKATKIPCESPPLEWVDTTASTKRLHRTVKOVQKEE 2736
QY 775 PTTIHKSPDESPELSABPTP-----KALENSKEPVGPTT-----810
Db 2737 PSAV-KFTQTSGETTDADKEPAGEDKGIKALKESAKQTPAPASVTGSSRRRPRAPRESAQ 2795
QY 811 -----KTPAA--TKPEMTTTAKDKTTERDLRTPE-----TTTAAFKMTKETATTTTEKT 857
Db 2796 ATIEDLAGFKDPAAGHTESMT--DDTKTKIPCKSSPELEDATSSKRRPRTEAQKVEVK 2852
QY 858 TE-----SKTATTQVTSITTO-----DTPFKITTLTKTLAPKVTTTKTITTTTEIM 907
Db 2853 EELLAVGKLTQTSGETTHTDKPVGEGKGTAKF-----QPAKRNVDAEDVI 2899
QY 908 NKPEETAFAKDKRA-----TNSKATTPKQKPTK--935
Db 2900 GSRQRPAPKEKAQPLEDLASFQELSQTGHTHEELANGAADSFSAKQTPDGGKPLKIS 2959
QY 936 --APK-KPT-----941
Db 2960 RRVLRAPKVEPVGDVVSTRDPVKSQSKSNTSLPPLPFKRGKGKGSVGTGKRLRCMPAPE 3019
QY 942 -----STKKPKTPRVKPKTTPT---PRKMTSTMPBLNPTSRIAEAMLOTTTRPNQ-990
Db 3020 EIVEELPASKQORVAPRARKGSSEPVVIMKRSLTRSAKIEPAEELNSDMKTNKEEHLK 3079
QY 991 --TNSKLVENVNPKSEDAGGAEGETHMLL-----RPHVPEPTEDDMDYLP 1035
Db 3080 QDSVENKGISRSRRQDKTEABQQITEVFVLAERIEINRNEKKP-----MKTSPEMD---3132
QY 1036 RVPNOGIILNP 1046
Db 3133 -----IQNP 3136
```

## RESULT 15

A41819

proline-rich peptides 637K precursor, prostatic - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 13-May-1992 #sequence\_revision 13-May-1992 #text\_change 09-Jul-2004

C:Accession: A42663; A41819; A31966; B20593; A20593

R:De Clercq, N.; Hemschoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.

J. Biol. Chem. 267, 9884-9894, 1992

A:Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the exon.

A:Reference number: A42663; MUID:92250652; PMID:1577819  
A:Accession: A42663  
A:Molecule type: DNA; mRNA  
A:Residues: 1-3706; 1', 3708-4077, 'F', 4079-4155, 'S', 4157-5761 <DE2>  
A:Cross-references: UNIPROT:Q63455; GB:M86514  
A:Experimental source: ventral prostate  
A>Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:100347, NCBI:100348)  
A:De Clercq, N.; Henschoote, K.; Devos, A.; Peters, B.; Heyns, W.; Rombauts, W.  
submitted to the Protein Sequence Database, April 1992  
A:Reference number: A41819  
A:Accession: A41819  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-5762 <DE1>  
R:Henschoote, K.; Peters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; Wind  
J. Biol. Chem. 263, 19159-19165, 1988  
A:Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich p  
A:Reference number: A31966; MUID:89066721; PMID:3198617  
A:Accession: A31966  
A:Molecule type: mRNA  
A:Residues: 3372-3540 <HEM>  
A:Cross-references: GB:M20721; GB:J04188; NID:9206397; PIDN:AAA41950.1; PID:G554494  
R:Peters, B.; Heyns, W.; Bossyns, D.; Rombauts, W.  
J. Biol. Chem. 258, 14206-14211, 1983  
A:Title: proline-rich polypeptides bound to rat prostatic binding protein. The primary s  
A:Reference number: A94675; MUID:84061859; PMID:6685733  
A:Accession: B20593  
A:Molecule type: protein  
A:Residues: 2020-2057 <PEE>  
A:Note: this peptide, designated proline-rich polypeptide V, can be found at several loc  
A:Accession: A20593  
A:Molecule type: protein  
A:Residues: 2822-2859 <PE2>  
A:Note: this peptide, designated proline-rich polypeptide IV, can be found at several loc  
C:Genetics:  
A:Introns: #status absent  
A:Note: single copy gene with no introns  
C:Superfamily: rat prostatic proline-rich peptides 637K precursor  
C:Keywords: prostate; tandem repeat  
F1-26/Domain: signal sequence #status predicted <SIG>  
F1:27-5761/Product: prostatic proline-rich peptides 637K proprotein #status predicted <MA

Query Match 14.8%; Score 826.9; DB 2; Length 5762;  
Best Local Similarity 13.5%; Pred. No. 7.9e-09;  
Matches 426; Conservative 151; Mismatches 411; Indels 2161; Gaps 116;

QY	4	KTLPIYLL	-----LLLSV-----	16
Db	456	ETLPIHEIQTQEDYGYQLPNVTVPVDVVALVTSEPVKTESFLAPQEPFVHALEYSN	515	
QY	17	----FVIOQ--VSSQELSKRCFCFSFERGRCDCDAQCKKYDK-----	54	
Db	516	DVEFFVNEEPPVQAPETGESQ-FES-----OLEVPAQATEYDEBFKTSATEQEQLAQF	569	
QY	55	-----	62	
Db	570	PENDEVTVLPSNHYQAQHSILSNVTQPLDLDTITEKPMEMGTSPPYVDAAAPEVEF	629	
QY	63	CAR-----BHSYSEN-----	72	
Db	630	LSQOQGVLSQSLPILYDLSLQCYETTGISQISEGGEPPFTQETPEHSVGMHTEVAQP	699	
QY	73	-----QSSSSSSSSSS	84	
Db	690	PGHVEVTGPPLGHGVHSPALONILTOYFTTFPEKQHSVPGLGVPGHLEEFSEVPSPSQ	749	
QY	85	SSTIW-----KIKSSKSAANRELQKLVKDKNKNRTKKXKPTPK	124	
Db	750	ENSAHVSVDMLFSLPVDLQIFRSTQSKSYKTIKHEDLALTI-----TPE	795	
QY	125	PPV-----VDEAGSG-----LDNGDFKVTTPDTS---	148	

Db	796	PSLEDGILFPQEDLLQPIDSTGQGFESHKSTLTSLKPPVSNKTSKSAFOETMSETTNS	855
QY	149	-----TTQNKVSTSP-----KIITAKP-----INP-----RPSL-----PP	175
Db	856	KQVDLSSTHLKSELPPDYTMGLEPSLYQOIITQSSPKSLHNPSPFVLKPSAAQALKPP	915
QY	176	NSDTSKETS-----LTVNKETTVETKETT-----	199
Db	916	RREKLSPTNNMIPHPSPKPLKNWTHIPAKHMTVPRIQEDQGEYIISNGSEFQPLDLEV	975
QY	200	-----TTNKQSTDGKEKTTSAKETQ-----STKTSKADLAPTSKVLAKP	240
Db	976	TLTSGIIPVKKHILPKRTVNPQTSQVKISHSHQVETQHPNSETTIVQPLDLLEFALNLP	1035
QY	241	TPK-----AETTTK-----GPAITTPKE-----	258
Db	1036	TPKENFAQTQFTTTOMIGP-----PREVIAQAPHEHGTIPPIQOAEYSTLPTVSFQP	1091
QY	259	-----PTPTTPKEPASTTPKE-----PTPT-----	278
Db	1092	LQDELITITSEALREPHPTVPQOTIIVHPPEHPLVIHSEQVYVQHNPTEALIQPLDLEL	1151
QY	279	TIKSAPTTPEKPAPTTKSAPTTPEKPAPTTKPEPAP-----TTTPE-----PAPT-	324
Db	1152	TITPQPTARGE-LPQTLQDSTTQIIEP-PTVVVGPVPIVEEVTQVTTSDQAEYPSPTV	1209
QY	325	-----TTTKE-----	328
Db	1210	SFQSLDELITISPEATRESYHPSLLQOTTIVNPPEHPLVIHSEQVYVQHNPTEALVQPL	1269
QY	329	-----PAPTPTTKSAPTTPK-----EPAPTTPKPAPTTP-----	357
Db	1270	DPDLTITPQTTTEGELPQTLQDSTSQIIEP-----PIKVVALVPVQVSDQAEYTTSS	1325
QY	358	-----KEP-APTTPE-----PTPT-----	371
Db	1326	TVSFQPLDQELTITSEALREPHPTVPQOTIIVHPPEHPLVIHSEQVYVQHNPTEALVQPL	1385
QY	372	-----TPKEPA-----PTTKEPAPTTPKEPAPTAPKPAPTTPKEPAPTTPKE-----PA	416
Db	1386	DLELWTPOPTARGELPQTLQDSTTQIIEP-PTVVVGPVPIVEEVTQVTTSDQAEYPPS	1444
QY	417	PTT-----TKEPSPT-----	430
Db	1445	PTVSFQSLDGLTITTEPTEHFTQKTTVPPPMYTDVTLPPQVSVQHLKTEGIVQPLDL	1504
QY	431	-----PAPT-----TTKSAPTTKEP-----APTTKSAPTT-----KEPSP	463
Db	1505	ELTITPQPTPEGBLSQTVQESTTQNKPEHVVAVPVVQAVTVPTPSQYAEYQKSLQP	1564
QY	464	-----TTTKEPA-----	472
Db	1565	LDELVTSEPTKEAVHSTIKSNLAINPQYVHIQHPNPAPAEATVQPLDLELTSSSLQPT	1624
QY	473	T-----PKE-----PAPTTPKKPAP-----	487
Db	1625	ABGELLYSMQETVTOISEPPKQVVPVPEYQEVAVPAPVDQAKYPLSSIVSLNSLDQEL	1684
QY	488	-----TTTKEPAPTTPK-----	499
Db	1685	TLSELLEGAHQUTTPDETWVLPKDRQGIYPDHDHKKHNLNTEVNTQPFLEHTVQHQP	1744
QY	500	-----EPAPTTPKKPAPTAPKEPAPTTP-----KETAPTTP-----	530
Db	1745	TIEERSQSIQKTTQITBPG-----KKVPLAQESEVETIPMILKETAFPTSHMALQS	1800
QY	531	-----KKLT-----	534
Db	1801	LDEKLAIHSHSPGWTQOAHNLKESKGTGKILLDYAEPNMEIELKHGFLKLTTEATT	1860
QY	535	-----	534
Db	1861	ESENTNQMTKSLKQVATLFTQNKKSMPLALVESQDSQPPPNMSLQPLDQELTSLSQPHG	1920

```

QY 535 -----PTTPEKL-----APTTEKPAPTT-----PEELAPTTP 562
Db 1921 WPHPIENTEKIYLHYAEPPTGPFVEPPELFFLKTKSRVQQTATMAASPKEMVSRAP 1980
QY 563 EE-----PTPTPEEPAP-----TPKAAAPNTPK-----EPA 590
Db 1981 ENKEAVLSGFGEDQDESPPNNMSLQDLQELTSSQPHGWIHPHNTHGKIYLHYAEP- 2039
QY 591 PTPP-----KEPAPTPK----- 603
Db 2040 PTGPFVEPPELFFLKTKSKPVEWTLTRDTSKSKEMVSPKYEAEAVLPVHGGEQESRS 2099
QY 604 -----EPAPTPP-----KETAP 615
Db 2100 PPNMSLQSLQELTSSQPHGVPHPNTHGKIYLHYAEP-PTGPFVEPPELFFLKTKS 2158
QY 616 TTPKGTAPTTLKEP----- 629
Db 2159 KPVQGTATRMVKSPEEMVSLDPENKEAVFPAQGEKGESPPNNMSLQSLDHLELFWSSQP 2218
QY 630 ---APTTPKKP-----APKELAPT 645
Db 2219 HGWIHPHNTPKIYLHYAEPPTGPFVEPPELFFLKTKSKPVQGTTEMAKSPKEMVSQ 2278
QY 646 TTK----- 648
Db 2279 TPYEKAVLSGPGEDQDESPPNTSLKSLDQEVAMSSQPHGVPHPKTPGKIYLHSIE 2338
QY 649 -----EST-----STTSCLKAPATPK----- 664
Db 2339 PPPGPFVKPTDLILVTKTKSPAETPRRIDKLKEMVPHSPYEBAVFPAGHGEQDES 2398
QY 665 GTAPTTPKEPA-----PTTP-----KEPAPTPKGTAP-----TT 694
Db 2399 GSPNNMPLQDLQELTSSQPHGVPHPNTHGKIYLHYAEP-PTGPFVEPPELFFLKTT 2457
QY 695 LKEPAPTTPKPA--PKELA----- 712
Db 2458 KSKPVQSGPKIAKSPKEMVSTPYEKAVLSQPGEDQDESPPNNMSLSLQDOVTMS 2517
QY 713 -----PTTKG-----PT-----STTSCLKAPATTP-----KET 735
Db 2518 QPHSGVPHPPKTPGKIYLHSIBPPPGPFVKPTDLILVTKTKSPAETPRRIDKLKEM 2577
QY 736 APTTPE-----PA-----PTTPKKP-----APTTPETP----- 759
Db 2578 VPHSPYEBAVFPAGGEQDESPPNNMPLQDLQELTSSQPHGVPHPNTHGKIYLH 2637
QY 760 ---PTTTSVSTP-----TTTKE-----PTTHKSPDE----- 784
Db 2638 YAEPTGPFVEPPELFFLKTKSKPVQGTQMAKSPPEMVSLSPKNETVFPQAQKGQD 2697
QY 785 ---STPELS-----AP----- 793
Db 2698 EGSPPNNISLQDLQELTSSQPHGWIHPHNTHGKIYLHYAEPPTGPFVEPPELFFLK 2757
QY 794 ---TPKALENSPKE----- 804
Db 2758 TSKSPMQSGPROIDKSPKEMFTQSPYEESLIPAHABGEQESRAPPHLSLQDLQELS 2817
QY 805 -----PGVP-----TTKPAATKP-EMTTAKDKT 828
Db 2818 SHPHGWIHPHNTPKIYLHYAEPPTGPFVEPPELFFLKTK-----SKPVQGTATKDKS 2873
QY 829 TERDLTTPETTTAA-----PKMT-----KETATTTT----- 855
Db 2874 PEDRVSTPYEKAVLSGPGEDQDESPPNNMSLQSLDQELALISSQPHGWIHPSPNAPDK 2933
QY 856 -----KTESK-ITATTQVSTTQ----- 875
Db 2934 IYLHYAEPPTGPFVEPPELFFLKTKSKPLOGTPTQMAKSPKEMVSTPYEKAELSAPG 2993

```

```

QY 876 -----DTTPF-----KIT 893
Db 2994 ENRDESPSSNNMSLHLDQELSSQPHGWIHPHNTHGKIYLHYAEPPTGPFVEPPEL 3053
QY 884 TLKTTTTLAPKVTTTKTITTTIMNKPEETAK--PKD----- 918
Db 3054 FLKTTTKSKP-----VQGSFQIDSKSPKEVFSQSPSESVLPAQAEQGEESR 3100
QY 919 -----RATNS 923
Db 3101 APFHMSLQDLQDLTSSHPHGWIHPHNTPGKIYLHYAEPPTGPFVEPPELFFLKTKS 3160
QY 924 KATTPKPKPTKAPK----- 939
Db 3161 KPVQSGPROVDKSPKEMFTQSPYEESVLPAAQAEQESRAPPHMSLQDLQDLTSSHP 3220
QY 940 -----PTSTKK-----PKIMPRVRKP-----KITPT-----PRMTST 967
Db 3221 HGWIHPHNTPKIYLHYAEPPTGPFVEPPELFFLKTKSKPAQWTPTQMAKSPPEMVSL 3280
QY 968 MPE-----LNPTSRIAEAMLQTT-----RPN----- 989
Db 3281 SPENKETVFPQAQKGQESISPPHMSLQDLQDLTSSHPHGWIHPHNTHGKIYLHYAE 3340
QY 990 -----QTPNSKLVEVNP----- 1001
Db 3341 PTGPFVEPPELFFLKTKSKPVQVSPROIDKSPKEVFTQSPYEESVLPAAQAEQGEESR 3400
QY 1002 ----- 1003
Db 3401 APPHMSLQDLQDLTSSHPHGWIHPHNTHGKIYLHYAEPPTGPFVEPPELFFLKTKS 3460
QY 1004 EDAGG-----AEGE-----TPHMLLR-----P 1020
Db 3461 KPVHSGPROIDKSKHEMFTQSPYEESVLPAAQAEQESRAPPHMSLQDLQDLTSSHP 3520
QY 1021 HVFMPE--VTPDMDYLPV-ENQGIINP 1046
Db 3521 HGWIHPHNTPKIYLHYAEPPTGPFVEP 3549

```

Search completed: October 13, 2004, 11:58:55

Job time : 65.3671 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 81.1065 Seconds  
(without alignments)  
4449.477 Million cell updates/sec

Title: SEQ1-E  
Perfect score: 5373  
Sequence: 1 MAWKTLPIVILLLLSVFVIQ.....DMDYLRVPRVQGIINPMLS 1006

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:.\*  
1: Geneseq1980s:.\*  
2: Geneseq1990s:.\*  
3: Geneseq2000s:.\*  
4: Geneseq2001s:.\*  
5: Geneseq2002s:.\*  
6: Geneseq2003as:.\*  
7: Geneseq2003bs:.\*  
8: Geneseq2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Match	Length	DB ID	Description
1	5373	100.0	1270	8	ADK67911	Adk67911 Human ext
2	5358.9	99.7	1311	8	ADK67912	Adk67912 Human ext
3	5349	99.6	1320	7	ADK65819	Adk65819 Angiogene
4	5339.6	99.4	1299	4	AAW24322	Aam24322 Human EST
5	5339.6	99.4	1404	2	AAW26049	Aar26049 MSF precu
6	5339.6	99.4	1404	4	AAW29773	Aab29773 Human meg
7	5339.6	99.4	1404	4	AAW60568	Aab60568 Human meg
8	5339.6	99.4	1404	8	ADM98014	Adm98014 Human meg
9	5339.6	99.4	1415	4	AAU32262	Aau32262 Novel hum
10	5330.6	99.2	1404	7	ADK65839	Adk65839 Angiogene
11	2920	54.3	546	4	ABU53252	Abu53252 Human tes
12	2888	53.8	551	4	ABU53253	Abu53253 Human tes
13	2880.1	53.6	902	4	AAW29778	Aab29778 Human MSF
14	2757	51.3	513	4	ABU53254	Abu53254 Human tes
15	1694.8	31.5	452	2	AAW80041	Aar80041 Human meg
16	1409.9	26.2	538	5	AAO18834	Aao18834 3' cartil
17	1335.3	24.9	5179	4	AAW24516	Aam24516 C899P pre
18	1335.3	24.9	5179	6	AAW55365	Abp55365 Human col
19	1335.3	24.9	5179	6	ABO07258	Abp07258 Human p53
20	1335.3	24.9	5179	7	ADD48091	Add48091 Human Pro
21	1335.3	24.9	5179	7	ADD44998	Add44998 Human Pro
22	1146.1	21.3	292	5	AAU11261	Aau11261 Human HAP
23	1108.6	20.6	1664	2	AAW43106	Aaw43106 C. thermo
24	1099.6	20.5	717	4	ABU53144	Abu53144 Human tes
25	1097	20.4	214	4	ABU53255	Abu53255 Human tes

## ALIGNMENTS

## RESULT 1

ADK67911  
ID ADK67911 standard; protein; 1270 AA.

XX AC ADK67911;

XX DT 06-MAY-2004 (first entry)

XX DE Human extracellular messenger (EXMES) polypeptide.

KW Human; extracellular messenger; EXMES; respiratory-gen.; anti-allergic;  
KW antiasthmatic; antiinflammatory; antidiabetic; neuroprotective;  
KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antiparasitic;  
KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;  
KW cytotstatic; gene therapy.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX FT Peptide 1..29

XX FT /label= Signal peptide  
XX FT /note= "Spans residues 1 to 18, 20, 21, 24, 27 or 29 according to identification method"

XX FT W02004013292-A2.

XX PD 12-FEB-2004.

XX XX 30-JUL-2003; 2003WO-US024084.

XX PR 02-AUG-2002; 2002US-0400810P.

XX PR 19-SEP-2002; 2002US-0412197P.

XX PR 04-OCT-2002; 2002US-0416004P.

XX PR 08-NOV-2002; 2002US-0424862P.

XX (INCY-) INCYTE CORP.

XX PI Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

XX PI Richardson TW, Emerling BM, Lindquist BA, Chawla NK, Ramkumar J;  
XX PI Lee SY;

XX DR WPI: 2004-157116/15.

XX XX N-PSDB; ADK67916.

XX PT New extracellular messengers and nucleic acids, useful for diagnosing,  
XX PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes  
XX PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or  
XX PT autoimmune thyroiditis.

26 1075.7 20.0 8991 6 ABU08487 Abu08487 S. pneumo  
27 1047.1 19.5 763 3 AAG38942 Aag38942 Arabidops  
28 1041.7 19.4 1795 4 ABB69806 Abb69806 Drosophil  
29 1039.7 19.4 1049 4 ABB61364 Abb61364 Drosophil  
30 1008.8 18.8 2284 4 ABB71434 Abb71434 Drosophil  
31 987.9 18.4 770 4 ABW53141 Abw53141 Human tes  
32 981 18.3 778 4 ABUS3143 Abu53143 Human tes  
33 974.4 18.1 692 4 ABUS3155 Abu53155 Human tes  
34 964.1 17.9 717 4 ABUS3150 Abu53150 Human tes  
35 964.1 17.9 717 4 ABUS3149 Abu53149 Human tes  
36 964.1 17.9 717 4 ABUS3151 Abu53151 Human tes  
37 964.1 17.9 717 4 ABUS3145 Abu53145 Human tes  
38 964.1 17.9 717 4 ABUS3148 Abu53148 Human tes  
39 964.1 17.9 717 4 ABUS3147 Abu53147 Human tes  
40 964.1 17.9 717 4 ABUS3146 Abu53146 Human tes  
41 955.4 17.8 5703 8 ADL23265 Adl23265 Human MUC  
42 952.1 17.7 745 4 ABUS3142 Abu53142 Human tes  
43 950.4 17.7 4315 5 ABP43908 Abp43908 MUC5B par  
44 948.6 17.7 2112 4 ABB60403 Abb60403 Drosophil  
45 932.7 17.4 1538 7 ADI21202 Adi21202 Novel hum

XX PS Claim 59; SEQ ID NO 4; 165pp; English.

XX CC The present sequence is that of novel human extracellular messenger

XX CC (EXMES) Incyte ID NO: 7513017CDI polypeptide. The protein shows homology

CC to human megakaryocyte stimulating factor. The invention provides EXMES

CC polynucleotides and polypeptides, as well as expression vectors, host

CC cells, antibodies, agonists and antagonists, and methods for diagnosing,

CC treating or preventing disorders associated with aberrant expression of

CC EXMES, especially autoimmune and inflammatory disorders, cell

CC proliferative disorders and endocrine disorders, e.g. adult respiratory

CC distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's

CC disease, diabetes mellitus, myasthenia gravis, osteoarthritis,

CC osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal,

CC parasitic, protozoal or helminthic infections, cancers, autoimmune

CC thyroiditis, cretinism, Plummer's disease or thyroid carcinoma.

CC Embodiments also provide methods for using the purified EXMES and/or

CC their encoding polynucleotides for facilitating the drug discovery

CC process, including determining of efficacy, dosage, toxicity and

CC pharmacology, and for investigating the pathogenesis of diseases and

CC medical conditions.

XX SQ Sequence 1270 AA;

Query Match 100.0%; Score 5373; DB 8; Length 1270;

Best Local Similarity 100.0%; Pred. No. 2.4e-142;

Matches 1006; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRGCFESFERGECDCDAQCKKYDKCCPDYE 60

DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRGCFESFERGECDCDAQCKKYDKCCPDYE 60

QY 61 SFCAEVKDNKNTKKTKPKPPVVDAGSLDNGDFKVTPTDSTTQHNVKSTSPKITT 120

DB 61 SFCAEVKDNKNTKKTKPKPPVVDAGSLDNGDFKVTPTDSTTQHNVKSTSPKITT 120

QY 121 AKDINRPSLPNNSDTSKETS LVNKEVTVETKTTNNKQSTDGKEKITS AKETQSIE 180

DB 121 AKDINRPSLPNNSDTSKETS LVNKEVTVETKTTNNKQSTDGKEKITS AKETQSIE 180

QY 181 KTSAKOLAPTSKVIAKTPKAEITTKGPALTTPKEPTTPPKEPASTTPKEPTTIKSA 240

DB 181 KTSAKOLAPTSKVIAKTPKAEITTKGPALTTPKEPTTPPKEPASTTPKEPTTIKSA 240

QY 241 PTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTKPE 300

DB 241 PTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTKPE 300

QY 301 PAPTTPKKPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 360

DB 301 PAPTTPKKPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTT 360

QY 361 PKEPAPTTKPEPAPTTTKESPTTPKPEPAPTTTKSAPTTTKSAPTTTKSAPTT 420

DB 361 PKEPAPTTKPEPAPTTTKESPTTPKPEPAPTTTKSAPTTTKSAPTTTKSAPTT 420

QY 421 TTTKPEPAPTTKPEPAPTTKPKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 480

DB 421 TTTKPEPAPTTKPEPAPTTKPKAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPE 480

QY 481 ETAPTTPKLPTTPPEKLAFTTPKPEPAPTTPEELAPTTPEEPTTPPEEPAPTT 540

DB 481 ETAPTTPKLPTTPPEKLAFTTPKPEPAPTTPEELAPTTPEEPTTPPEEPAPTT 540

QY 541 NTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTK 600

DB 541 NTPKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTKPEPAPTTK 600

QY 601 PTTTKETSTSDKPEAPTTKGTAPTTKPEPAPTTKPEPAPTTKGTAPTTLKPEPAP 660

DB 601 PTTTKETSTSDKPEAPTTKGTAPTTKPEPAPTTKPEPAPTTKGTAPTTLKPEPAP 660

QY 661 KKPAPKELAPTTTKGPTSTTSDKPAFTTPKETAFTTPKPEPAPTTPKKPAFTT 720

DB 661 KKPAPKELAPTTTKGPTSTTSDKPAFTTPKETAFTTPKPEPAPTTPKKPAFTT 720

QY 721 SEVSTPTTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKPGVPTTKTPAATKPEMTT 780

DB 721 SEVSTPTTTTKEPTTIHKSPDESTPELSAEPTPKALENSPKPGVPTTKTPAATKPEMTT 780

QY 781 AKDKTTERDLRTPETTTAAPKWKETATTBKTTESKITATTQVTSITQDTPPKIT 840

DB 781 AKDKTTERDLRTPETTTAAPKWKETATTBKTTESKITATTQVTSITQDTPPKIT 840

QY 841 TLKTTTLAPKVTITTKKTIITTTIMNKPEETAPEAKPKORATNSKATTPKPKKPTST 900

DB 841 TLKTTTLAPKVTITTKKTIITTTIMNKPEETAPEAKPKORATNSKATTPKPKKPTST 900

QY 901 KKPKNMPVRVKPTTTPPKMTSTMPELNPTSRIBAMLOTTTRPNQOTNSKLVEVNP 960

DB 901 KKPKNMPVRVKPTTTPPKMTSTMPELNPTSRIBAMLOTTTRPNQOTNSKLVEVNP 960

QY 961 EDAGGAGETGPHMLLRPHVFMPEVTPDMYLPKVPNQGIIINPMLS 1006

DB 961 EDAGGAGETGPHMLLRPHVFMPEVTPDMYLPKVPNQGIIINPMLS 1006

RESULT 2

ADK67912

ID ADK67912 standard; protein; 1311 AA.

XX AC ADK67912;

XX DT 06-MAY-2004 (first entry)

XX DE Human extracellular messenger (EXMES) polypeptide.

XX KW Human; extracellular messenger; EXMES; respiratory-gen.; anti-allergic;

KW antiasthmatic; anti-inflammatory; antidiabetic; neuroprotective;

KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;

KW virucide; fungicide; antiparasitic; protozoacide; antihelminthic;

KW cytostatic; gene therapy.

XX OS Homo sapiens.

XX FH Key

FT Peptide

FT 1. .30

FT /label= Signal\_peptide

FT /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30

FT according to identification method"

XX WO2004013292-A2.

XX PD 12-FEB-2004.

XX PF 30-JUL-2003; 2003WO-US024084.

XX PR 02-AUG-2002; 2002US-0400810P.

PR 19-SEP-2002; 2002US-0412197P.

PR 04-OCT-2002; 2002US-0416004P.

PR 08-NOV-2002; 2002US-0424862P.

XX (INCY-) INCYTE CORP.

XX PI Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;

PI Richardson JW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J;

PI Lee SY;

XX WPI; 2004-157116/15.

DR N-PSDB; ADK67917.

XX New extracellular messengers and nucleic acids, useful for diagnosing,

PT treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes

PT mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or

PT autoimmune thyroiditis.

XX



Claim 60; SEQ ID NO 5; 165pp; English.  
 The present sequence is that of novel human extracellular messenger (EXMES) Incyte ID NO: 7513018CD1 polypeptide. The protein shows homology to human megakaryocyte stimulating factor. The invention provides EXMES polynucleotides and polypeptides, as well as expression vectors, host cells, antibodies, agonists and antagonists, and methods for diagnosing, treating or preventing disorders associated with aberrant expression of EXMES, especially autoimmune and inflammatory disorders, cell proliferative disorders and endocrine disorders, e.g. adult respiratory distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, parasitic, protozoal or helminthic infections, cancers, autoimmune thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. Embodiments also provide methods for using the purified EXMES and/or their encoding polynucleotides for facilitating the drug discovery process, including determining of efficacy, dosage, toxicity and pharmacology, and for investigating the pathogenesis of diseases and medical conditions.

Query Match 99.7%; Score 5358.9; DB 8; Length 1311;  
 Best Local Similarity 96.1%; Pred. No. 6.2e-142;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

Qy	1	MAWKTLPIYLILLLSVFIQVSSQ-----	----- 25
Db	1	MAWKTLPIYLILLLSVFIQVSSQDLSSACRGCGYSRDATCNCYDNCQHYMECCPDF	60
Qy	26	-----ELSCKGRCFESFERGREGCDCAQCKKYDKCCPDYESFCAEVKDNKKNTKKKPT	79
Db	61	KRVCTAELSCKGRCFESFERGREGCDCAQCKKYDKCCPDYESFCAEVKDNKKNTKKKPT	120
Qy	80	PKPVPVDEAGSLDNGDFKVTTPDTSTTOHNKVSTSPKIIITAKINPRPSLPNSDTSKE	139
Db	121	PKPVPVDEAGSLDNGDFKVTTPDTSTTOHNKVSTSPKIIITAKINPRPSLPNSDTSKE	180
Qy	140	TSLVNKETVETVEKETTITNNKQSTDCGKEKITSASKEQTSEKTSAKOLAPTSKVLAKPTP	199
Db	181	TSLVNKETVETVEKETTITNNKQSTDCGKEKITSASKEQTSEKTSAKOLAPTSKVLAKPTP	240
Qy	200	KAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTKSAPTTP	259
Db	241	KAETTTKGPALTTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTKSAPTTP	300
Qy	260	KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTKEPAPTTPKGPAPTTKEPAP	319
Db	301	KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKSAPTTTKEPAPTTPKGPAPTTKEPAP	360
Qy	320	TTPEKPTPTTPKEPAPTTKEPAPTTKEPAPTA PKPAPTTTKEPAPTTTKEPAPTTTKE	379
Db	361	TTPEKPTPTTPKEPAPTTKEPAPTTKEPAPTA PKPAPTTTKEPAPTTTKEPAPTTTKE	420
Qy	380	PSPTTKEPAPTTTKSAPTTTKGPAPTTTKSAPTTTKEPSPTTTKEPAPTTTKEPAPTTTP	439
Db	421	PSPTTKEPAPTTTKSAPTTTKGPAPTTTKSAPTTTKEPSPTTTKEPAPTTTKEPAPTTTP	480
Qy	440	KKPAPTTKEPAPTTTKEPAPTTTKKPAPTA PKPAPTTTKEPAPTTTKEPAPTTTKEPAP	499
Db	481	KKPAPTTTKEPAPTTTKEPAPTTTKKPAPTA PKPAPTTTKEPAPTTTKEPAPTTTKEPAP	540
Qy	500	PTTPEKAPPTTPEELAPTTPEEPTTTPPEPAPTTTPKAAAPNTPKEPAPTTTKEPAPTTTP	559
Db	541	PTTPEKAPPTTPEELAPTTPEEPTTTPPEPAPTTTPKAAAPNTPKEPAPTTTKEPAPTTTP	600
Qy	560	KEPAPTTPKETAPTTPKGTAPTTTKGPAPTTTPKKPAPKELAPTTTKEPTSTTSDKKAPTT	619
Db	601	KEPAPTTPKETAPTTPKGTAPTTTKGPAPTTTPKKPAPKELAPTTTKEPTSTTSDKKAPTT	660
Qy	620	PKGTAPTTKEPAPTTTKEPAPTTTPKGTAPTTTKGPAPTTTPKKPAPKELAPTTTKEPTST	679

CC the angiogenic index. The methods and compositions of the present  
CC invention are useful for diagnosing, preventing and/or treating cancer,  
CC coronary artery disease, myocardial ischemia or coronary  
CC arteriosclerosis. They can also be used in research, drug discovery and  
CC forensic medicine involving angiogenesis. This sequence corresponds to  
CC one of the differentially expressed proteins of the invention.  
XX  
SQ Sequence 1320 AA;  
Query Match 99.6%; Score 5349; DB 7; Length 1320;  
Best Local Similarity 95.1%; Pred. No. 1.2e-141;  
Matches 1004; Conservative 0; Mismatches 2; Indels 50; Gaps 1;  
QY 1 MAWKTLPIYLLLLSVFVIQVSSQBELSKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
DB 1 MAWKTLPIYLLLLSVFVIQVSSQBELSKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
QY 61 SFCA-----EYKDNK 70  
DB 61 SFCAEHNTPSPSSKKARPPSGASQTIKSTKSPKPPNKKTKKVISEIEITEVKDNK 120  
QY 71 KNRKKKPTPKPPVDEAGSLDNGDFKVTPDTSTTQHNKVSTSPKITTAKPINRPSL 130  
DB 121 KNRKKKPTPKPPVDEAGSLDNGDFKVTPDTSTTQHNKVSTSPKITTAKPINRPSL 180  
QY 131 PMSDTSKETSILVNETVETKETTNNKQTSQTSKETSQTSIEKTSKADLAPT 190  
DB 181 PMSDTSKETSILVNETVETKETTNNKQTSQTSKETSQTSIEKTSKADLAPT 240  
QY 191 SKVLAKPTKAEITTKGPAITTKPEPTTPKPEASTTKPEPTTIKSAPTTPKEPAPT 250  
DB 241 SKVLAKPTKAEITTKGPAITTKPEPTTPKPEASTTKPEPTTIKSAPTTPKEPAPT 300  
QY 251 TTKSAPTTKEPAPTTTKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 310  
DB 301 TTKSAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 360  
QY 311 PTTKEPAPTTKEPTTTKEPAPTTKEPAPTTKEPAPTTAPKAPAPTTKEPAPTTPK 370  
DB 361 PTTKEPAPTTKEPTTTKEPAPTTKEPAPTTKEPAPTTAPKAPAPTTKEPAPTTPK 420  
QY 371 EPAPTTTKEPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 430  
DB 421 EPAPTTTKEPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 480  
QY 431 PKEPAPTTPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTAPKAPAPTTKEPAPTT 490  
DB 481 PKEPAPTTPKAPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 540  
QY 491 TPTTPEKLAAPTTKEPAPTTPEELAPTTPEPTTPEEPAPTTPKAAAPNTPKAPAPT 550  
DB 541 TPTTPEKLAAPTTKEPAPTTPEELAPTTPEPTTPEEPAPTTPKAAAPNTPKAPAPT 600  
QY 551 PKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 610  
DB 601 PKEPAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 660  
QY 611 TSDKAPATTGTAAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 670  
DB 661 TCDKAPATTGTAAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 720  
QY 671 TTTKGPSITSDKAPATTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 730  
DB 721 TTTKGPSITSDKAPATTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 780  
QY 731 EPTTIHKSPDESTPELSAETPKALENSPKPEPGVPTTKTAPATKPEMTTAKOKTTERDL 790  
DB 781 EPTTIHKSPDESTPELSAETPKALENSPKPEPGVPTTKTAPATKPEMTTAKOKTTERDL 840  
QY 791 RTTPETTTAAAPKMTKETATTTTEKTESKIRATTTTQVTSITTTQDTPPKITLTKTTLAPK 850  
DB 841 RTTPETTTAAAPKMTKETATTTTEKTESKIRATTTTQVTSITTTQDTPPKITLTKTTLAPK 900

QY 851 VTTTKTTTITTTIMNKPETAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKTNPRVR 910  
DB 901 VTTTKTTTITTTIMNKPETAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKTNPRVR 960  
QY 911 KFKTTTPTRKMTSTMPELNPTSRIAEAMLTQTTTRNQTNSKLVNPKSEDAAGAEGET 970  
DB 961 KFKTTTPTRKMTSTMPELNPTSRIAEAMLTQTTTRNQTNSKLVNPKSEDAAGAEGET 1020  
QY 971 PHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1006  
DB 1021 PHMLLRPHVFMPEVTPDMDYLPRVNOGIIINPMLS 1056  
RESULT 4  
ID AAM24322 standard; protein; 1299 AA.  
XX AAM24322;  
XX AC AAM24322;  
XX DT 12-OCT-2001 (first entry)  
XX DE Human EST encoded protein SEQ ID NO: 1847.  
XX KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
XX KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;  
XX KW gene therapy; nutrition.  
XX OS Homo sapiens.  
XX WO200154477-A2.  
XX PD 02-AUG-2001.  
XX PF 25-JAN-2001; 2001WO-US002687.  
XX PR 25-JAN-2000; 2000US-00491404.  
XX PR 17-JUL-2000; 2000US-00617746.  
XX PR 03-AUG-2000; 2000US-00631451.  
XX PR 15-SEP-2000; 2000US-00663870.  
XX (HYSE-) HYSEQ INC.  
XX PA Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
XX PI Cao Y, Drmanac RA, Zhang J, Wehrman T;  
XX DR WPI; 2001-476164/51.  
XX DR N-PSDB; AAH98981.  
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
XX antibodies and research use.  
XX PS Claim 20; Page 1198-1201; 1275pp; English.  
XX CC The present invention provides the protein and coding sequences of novel  
XX proteins from a variety of organisms, including human, dog, cat, horse,  
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)  
XX from the organism of interest. They can be used in diagnostics,  
XX forensics, gene mapping, identification of mutations, to assess  
XX biodiversity and for nutritional purposes. The present sequence is a  
XX protein of the invention  
SQ Sequence 1299 AA;  
Query Match 99.4%; Score 5339.6; DB 4; Length 1299;  
Best Local Similarity 88.2%; Pred. No. 2.1e-141;  
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;  
QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25  
DB 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGYSRDATCNCYDNCQHYMECCPDF 60

QY 26 -----ELSKGRCFESFERGECDCDAQCKYDKCCPDYESFCAE----- 65  
 Db |||||  
 61 KRVTAEELSKGRCFESFERGECDCDAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120  
 QY 66 ----- 65  
 Db |||||  
 121 PPSGASOTIKSTTKRSPKPNKXKTKVIESEEITEHSHSVSENOESSSSSSSSSTIW 180  
 QY 66 -----VKDNKKNTKKKPTPKPPVWDEAGSLONGDFKVTPTDTST 106  
 Db |||||  
 181 KIKSSKNSAANRELQKKLVKDNKNRKKKPTPKPPVWDEAGSLONGDFKVTPTDTST 240  
 QY 107 TOHNKIVSTSKITTAKEINPRPSLPNSDTSKETSITVNETKETTITNNKQTSIDG 166  
 Db 241 TOHNKIVSTSPKITTAKEINPRPSLPNSDTSKETSITVNETKETTITNNKQTSIDG 300  
 QY 167 KEKTTSAKETQSTIEKTSKAKOLAPTSKVLAKPTPKAETTTKGPALTTPKPTPTTKBPAS 226  
 Db |||||  
 301 KEKTTSAKETQSTIEKTSKAKOLAPTSKVLAKPTPKAETTTKGPALTTPKPTPTTKBPAS 360  
 QY 227 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 286  
 Db |||||  
 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 420  
 QY 287 APTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 346  
 Db |||||  
 421 APTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 480  
 QY 347 EPAPTAPKAPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 406  
 Db |||||  
 481 EPAPTAPKAPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 540  
 QY 407 TTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 466  
 Db |||||  
 541 TTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 600  
 QY 467 APTAPKEPAPTTPKETAPTTPKLPTTPPKLAPTTPEKAPTTPEELAPTTPEEPTPTT 526  
 Db |||||  
 601 APTAPKEPAPTTPKETAPTTPKLPTTPPKLAPTTPEKAPTTPEELAPTTPEEPTPTT 660  
 QY 527 PREPAPTTPKAAAPNTPKAPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 586  
 Db |||||  
 661 PREPAPTTPKAAAPNTPKAPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEP 720  
 QY 587 APTTPKAPAPKELAPTTTKETPTSTSDKAPAPTTPKGAPTTTPKEPAPTTTPKEPAPTTTKEP 646  
 Db |||||  
 721 APTTPKAPAPKELAPTTTKETPTSTSDKAPAPTTPKGAPTTTPKEPAPTTTPKEPAPTTTKEP 780  
 QY 647 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPAPTTPKETAPTTTPKEPAPTTTKEP 706  
 Db |||||  
 781 TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTSDKAPAPTTPKETAPTTTPKEPAPTTTKEP 840  
 QY 707 KPAPTTPETPTPTTSEVSTPTTKETTTTHKSDSTPELSABPTPKALENSPKEPGVPT 766  
 Db |||||  
 841 KPAPTTPETPTPTTSEVSTPTTKETTTTHKSDSTPELSABPTPKALENSPKEPGVPT 900  
 QY 767 TKTPAATKPEMITTAKDKTTERDLRTTPTTTAAAPKMTKETAITTTKTTESKITATTTQV 826  
 Db |||||  
 901 TKTPAATKPEMITTAKDKTTERDLRTTPTTTAAAPKMTKETAITTTKTTESKITATTTQV 960  
 QY 827 TSTTTQDTPTPKITTLKTTTLAPKVTTKKITTTTEIMNKPEETAKPKDRATNSKATTPK 886  
 Db |||||  
 961 TSTTTQDTPTPKITTLKTTTLAPKVTTKKITTTTEIMNKPEETAKPKDRATNSKATTPK 1020  
 QY 887 POKPTKAPKPTSTKPKMTMPVRKKTITTPKMTSTMPNLNPTSRIAEAMLQTTTRN 946  
 Db |||||  
 1021 POKPTKAPKPTSTKPKMTMPVRKKTITTPKMTSTMPNLNPTSRIAEAMLQTTTRN 1080  
 QY 947 QTPNSKLVNPKSDEAGGAGETPHMLLRPHVMEPEVTPDMDYLPVNPQGIINPMLS 1006  
 Db 1081 QTPNSKLVNPKSDEAGGAGETPHMLLRPHVMEPEVTPDMDYLPVNPQGIINPMLS 1140

RESULT 5  
 AAR26049  
 ID AAR26049 standard; protein; 1404 AA.  
 XX AC AAR26049;  
 XX DT 25-MAR-2003 (revised)  
 DT 02-FEB-1993 (first entry)  
 XX MSF precursor.  
 DE Megakaryocyte colony stimulating factor; secretion signal; meg-CSF;  
 KW stability; proteolytic cleavage; adhesion; alternative splicing.  
 OS Synthetic.  
 XX  
 Key Location/Qualifiers  
 Region 1..26 /label= Exon\_I  
 Region 26..67 /label= Exon\_II  
 Region 67..107 /label= Exon\_III  
 Region 107..157 /label= Exon\_IV  
 Region 157..200 /label= Exon\_V  
 Region 200..1141 /label= Exon\_VI  
 Region 1166..1212 /label= Exon\_VII  
 Region 1213..1266 /label= Exon\_VIII  
 Region 1266..1331 /label= Exon\_IX  
 Region 1331..1373 /label= Exon\_X  
 Region 1373..1404 /label= Exon\_XI  
 Region 1411..1166 /label= Exon\_XII  
 Region /label= Exon\_VII  
 WC9213075-AL.  
 06-AUG-1992.  
 17-JAN-1992; 92WO-US000433.  
 18-JAN-1991; 91US-00643502.  
 10-SEP-1991; 91US-00757022.  
 (GEMY ) GENETICS INST INC.  
 Turner K, Clark SC, Jacobs K, Hewick RM, Gesner TG;  
 WPI; 1992-284660/34.  
 N-PSDB; AAQ27223.  
 New human mega-karyocyte stimulating factors - for treating immune  
 deficiencies, cancer, exposure to radiation or drugs, bacterial and viral  
 infections, etc.  
 Claim 1, 2 and 3; Fig 1; 87pp; English.  
 The sequence given is a full length translation from the megakaryocyte  
 stimulating factor (MSF) precursor. The sequence covered by exons I, III  
 and IV encodes megakaryocyte stimulating factor (MSF). This sequence is  
 modified by the addition of an N-terminal sequence encoding a secretory  
 leader, an initiating methionine preceding exon II and a terminating  
 codon following exon IV. The cDNA sequence given contains sequences  
 derived from human megakaryocyte colony stimulating factor (meg-CSF).  
 Exon I contains the initiating methionine, and encodes a classical  
 mammalian protein secretion signal sequence. The sequence encoding the

CC original meg-CSF includes exons II-IV and is thought to terminate in the  
CC region between amino acid residues 134 - 147. The primary transcript of  
CC this gene may be cleaved in different ways to yield a family of mRNA's  
CC each encoding a different MSF protein. Exons V and VI are thought to be  
CC related to the activity of the factor and are also implicated in the  
CC stability, folding and processing of the molecule. These exons are also  
CC thought to play a role in the observed synergy of MSF with other  
CC cytokines. Exons V - XII are believed to be implicated in the processing  
CC or folding of the appropriate structure of the resulting factor, ie. one  
CC or more of these exons may contain sequences which direct proteolytic  
CC cleavage, adhesion, organisation of the cellular matrix or extracellular  
CC matrix processing. Both naturally occurring and non-naturally occurring  
CC MSF's may be characterised by various combinations of alternatively  
CC spliced exons from this sequence, with the exons spliced together in  
CC differing orders to form different members of the MSF family. (Updated on  
CC 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 1404 AA;  
  
Query Match 99.4%; Score 5339.6; DB 2; Length 1404;  
Best Local Similarity 88.2%; Pred. No. 2.3e-141;  
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;  
  
QY 1 MAWKTLPIYLILLLSVFIQVSSQ----- 25  
DB 1 MAWKTLPIYLILLLSVFIQVSSQDLSSCAGRGEGYSRDATCNDYCNQHWECCPDF 60  
  
QY 26 -----ELSCKGRCFESFERGECDDCAQCKYDKCCPDYESFCAE----- 65  
DB 61 KRVCTAELSCKGRCFESFERGECDDCAQCKYDKCCPDYESFCAEVHNPTSPSSKKAP 120  
QY 66 ----- 65  
DB 121 PPSGASQTIKSTTRSPKPPNKKTKKXVIESEITEHSHSVSNQESSSSSSSSSTI 180  
QY 66 -----VKDNKNRTKKPKPPVDBAGSLDNGDFKVTTPDTST 106  
DB 181 KIKSSNKAANRELQKLVKVDNKNKNTKKPKPPVDBAGSLDNGDFKVTTPDTST 240  
QY 107 TOHNKYSTPKITTAQPINRPSLPNSDTSKETSLSLVNKEVVETKETTNNKQSTDG 166  
DB 241 TOHNKYSTPKITTAQPINRPSLPNSDTSKETSLSLVNKEVVETKETTNNKQSTDG 300  
QY 167 KEKTTSAKETQSIKTSKADLAPTSKVIAKPTPKAETTKGPAITTKPEPTTPKEPAS 226  
DB 301 KEKTTSAKETQSIKTSKADLAPTSKVIAKPTPKAETTKGPAITTKPEPTTPKEPAS 360  
QY 227 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEP 286  
DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKEPAPTTTKEP 420  
QY 287 APTTKSAPTTPKEPAPTTPKKAPTTPKEPAPTTPKKEPTPTTPKEAPTTKAP 346  
DB 421 APTTKSAPTTPKEPAPTTPKKAPTTPKEPAPTTPKKEPTPTTPKEAPTTKAP 480  
QY 347 EPAPTAKKAPTTPKEPAPTTPKEPAPTTTKESPTTPKEAPTTTKSAPTTTKEP 406  
DB 481 EPAPTAKKAPTTPKEPAPTTPKEPAPTTTKESPTTPKEAPTTTKSAPTTTKEP 540  
QY 407 TTKSAPTTPKPSPTTTTKEPAPTTPKEPAPTTPKKAPTTPKEPAPTTTKK 466  
DB 541 TTKSAPTTPKPSPTTTTKEPAPTTPKEPAPTTPKKAPTTPKEPAPTTTKK 600  
QY 467 APTAPKGPAPTTPKETAPTTPKKLTPTTTPPEKLAOTTPPEKAPTTPEELAPTTPEPTPT 526  
DB 601 APTAPKGPAPTTPKETAPTTPKKLTPTTTPPEKLAOTTPPEKAPTTPEELAPTTPEPTPT 660  
QY 527 PEEEPAPTTKAAANTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 586  
DB 661 PEEEPAPTTKAAANTPKEPAPTTPKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEP 720  
QY 587 APTTPKPKAPKELAPTTTKEPTSTSDKPAETTPKGTAPTTPKPAETTPKPAETTPKG 646

DB 721 APTTPKPKAPKELAPTTTKEPTSTSDKPAETTPKGTAPTTPKPAETTPKPAETTPKG 780  
QY 647 TAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKPAETTPKGTAPTTPKPAETTPK 706  
DB 781 TAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTSDKPAETTPKGTAPTTPKPAETTPK 840  
QY 707 KEAPTTPEPTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETTPKALENSKPEGVPT 766  
DB 841 KEAPTTPEPTPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETTPKALENSKPEGVPT 900  
QY 767 TKTPAATKPEMTTAKDKTTERDLRTPPTTTAAPKMTKETAATTTKTTESKITATTTQV 826  
DB 901 TKTPAATKPEMTTAKDKTTERDLRTPPTTTAAPKMTKETAATTTKTTESKITATTTQV 960  
QY 827 TSTTTQDTTTFPKITTLTKTTLAPKVTTKKTIITTEIMNKPEBTAKPKDRATNSKATTPK 886  
DB 961 TSTTTQDTTTFPKITTLTKTTLAPKVTTKKTIITTEIMNKPEBTAKPKDRATNSKATTPK 1020  
QY 887 PQKPTKAPKKPTSTKKPKTMPRVKPKTTTTPKMTSTMPELNPTSRISAEAMLOTTTRPN 946  
DB 1021 PQKPTKAPKKPTSTKKPKTMPRVKPKTTTTPKMTSTMPELNPTSRISAEAMLOTTTRPN 1080  
QY 947 QTPNSKLVEVNPVKSEDAAGAGETPHMLLRPHVFMPEVTPDMYDLPRVFNQGIINPMLS 1006  
DB 1081 QTPNSKLVEVNPVKSEDAAGAGETPHMLLRPHVFMPEVTPDMYDLPRVFNQGIINPMLS 1140  
  
RESULT 6  
AAB29773  
ID AAB29773 standard; protein; 1404 AA.  
AC AAB29773;  
XX  
DT 28-FEB-2001 (first entry)  
XX Human megakaryocyte stimulating factor (MSF), SEQ ID NO:1.  
DE  
XX Human MSF; megakaryocyte stimulating factor; tribonectin;  
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;  
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;  
KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.  
XX  
OS Homo sapiens.  
XX  
XX W0200064930-A2.  
PN  
XX  
PD 02-NOV-2000.  
XX  
XX 24-APR-2000; 2000WO-US010953.  
PF  
XX  
XX 23-APR-1999; 99US-00298970.  
PR  
XX (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.  
PA  
XX Jay GD;  
PI  
XX  
XX WPI: 2001-024673/03.  
DR  
XX N-PSDB; AAC81498.  
DR  
XX Novel tribonectin polypeptide useful as lubricant for treating  
PT osteoarthritis, comprises O-linked lubricating moiety.  
PT  
XX  
PS Claim 3; Page 7; 47pp; English.  
XX  
XX The invention relates to a human tribonectin which is a product of  
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)  
CC gene. The tribonectin has at least one O-linked oligosaccharide  
CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats  
CC of a motif having at least 50% identity to the sequence KEAPTT  
CC (AAB29774). The invention also relates to a nucleic acid encoding a human  
CC MSF-derived tribonectin; a biocompatible composition comprising a human  
CC tribonectin for inhibiting tissue adhesion formation; and a method of  
CC diagnosing osteoarthritis or a predisposition to osteoarthritis by

measuring the amount of MSF or its fragment in a biological sample of a mammal, wherein an increased amount of MSF compared to a control indicates the presence of or predisposition to developing osteoarthritis. The tribonectin and DNA encoding it are useful in the treatment of osteoarthritis, where they may be used for lubricating mammalian joints, such as articulating joints of humans, dogs or horses. The tribonectin, when formulated as a membrane, foam, gel or fibre, is useful for inhibiting adhesion between two surfaces such as the injured tissues of a mammal, where the injury is caused by a surgical insertion or trauma, or an artificial device e.g., an orthopaedic implant. In particular, one of the surfaces is pericardial tissue. DNA encoding a tribonectin may be used in gene therapy. The present sequence represents human MSF

Seq Sequence 1404 AA;

Query Match 99.4%; Score 5339.6; DB 4; Length 1404;  
 Best Local Similarity 88.2%; Pred. No. 2.3e-141;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLSVFIQVSSQ----- 25  
 DB 1 MAWKTLPIYLLLSVFIQVSSQDLSSCAGRCGEGSRDATCNCDCYNQHYMECCPDF 60

QY 26 -----ELCKGRCFSPRGRECDCAOQCKYDKCCPDYESFCAE----- 65  
 DB 61 KRVCTAELCKGRCFSPRGRECDCAOQCKYDKCCPDYESFCAEVHNPSTSPSSKKAP 120

QY 66 ----- 65

DB 121 PPSGASQTIKTKRSPKPPNKKTKVIESEBITTEHSVSENQESSSSSSSSSTIW 180

QY 66 -----VKONKNRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST 106  
 DB 181 KIKSKNSAANRELQKLVKONKNRTKKKPPKPPVVDVDEAGSLDNGDFKVTTPDTST 240

QY 107 TQHNKVSIPKITTAKPINRPSLPNSDTSKETSUTVNKETTVEPKETTTNNKQSTDG 166  
 DB 241 TQHNKVSIPKITTAKPINRPSLPNSDTSKETSUTVNKETTVEPKETTTNNKQSTDG 300

QY 167 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGNALTPKPEPTTTKPEPAS 226  
 DB 301 KEKTTSAKETQSIKTSADLAPTSKVLAKPTPKAETTTKGNALTPKPEPTTTKPEPAS 360

QY 227 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTKEP 286  
 DB 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTPKEPAPTTTKEP 420

QY 287 APTTTKSAPTTPKEPAPTTPKPAPTTTPKEPAPTTPKETPTTPKEPAPTTTPK 346  
 DB 421 APTTTKSAPTTPKEPAPTTPKPAPTTTPKEPAPTTPKETPTTPKEPAPTTTPK 480

QY 347 EPAPTAPKPPAPTTPKEPAPTTTPKEPAPTTPKKAPTTTPKEPAPTTTPK 406  
 DB 481 EPAPTAPKPPAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 540

QY 407 TTKSAPTTPKPSPTTTKEPAPTTTPKEPAPTTPKKAPTTTPKEPAPTTTPK 466  
 DB 541 TTKSAPTTPKPSPTTTKEPAPTTTPKEPAPTTPKKAPTTTPKEPAPTTTPK 600

QY 467 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPKEPAPTTPEELAPTTPEEPTPT 526  
 DB 601 APTAPKEPAPTTPKETAPTTPKLTPTTPEKLAPTTPKEPAPTTPEELAPTTPEEPTPT 660

QY 527 PREPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 586  
 DB 661 PREPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 720

QY 587 APTTPKAPKELAPTTTKETSTSDKPAPTTPKGTAPTTPKEPAPTTTPK 646  
 DB 721 APTTPKAPKELAPTTTKETSTSDKPAPTTPKGTAPTTPKEPAPTTTPK 780

QY 647 TAPTTLKPEPAPTTPKKAPKELAPTTTKGPTSTTSKPAPTTPKETAPTTPK 706

DB 781 TAPTTLKPEPAPTTPKKAPKELAPTTTKGPTSTTSKPAPTTPKETAPTTPK 840  
 QY 707 KPAPTTPPEPTTPPTTSEVSTPTTTKEPTTIHKSDESTPELSABPTPKALENSPK 766  
 DB 841 KPAPTTPPEPTTPPTTSEVSTPTTTKEPTTIHKSDESTPELSABPTPKALENSPK 900  
 QY 767 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTESKITATTQV 826  
 DB 901 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETATTTESKITATTQV 960  
 QY 827 TSTTTQDTTPFKITTLTKTTTLAPKVTTKKTIITTEIMNKPEETAKEPKORATNSKATTPK 886  
 DB 961 TSTTTQDTTPFKITTLTKTTTLAPKVTTKKTIITTEIMNKPEETAKEPKORATNSKATTPK 1020  
 QY 887 PQKPTKAPKPTSTKXPKTMVRVRKPKTTTPPKXMTSTMPELNPTSRIAEAMLOTTTRN 946  
 DB 1021 PQKPTKAPKPTSTKXPKTMVRVRKPKTTTPPKXMTSTMPELNPTSRIAEAMLOTTTRN 1080

QY 947 QTPNSKLVVNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1006  
 DB 1081 QTPNSKLVVNPKSDEAGGAGETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1140

RESULT 7  
 AAB60568  
 ID AAB60568 standard; protein; 1404 AA.  
 XX AC AAB60568;  
 XX DT 27-APR-2001 (first entry)  
 XX DE Human megakaryocyte stimulating factor (MSF, CACP).  
 XX KW Human; CACP protein; camptodactyly-arthropathy-coxa vara-pericarditis;  
 XX KW MSF; megakaryocyte stimulating factor; synovial lubricant;  
 XX KW chromosome 1q25-31; osteoarthritis; joint lubrication; osteopathic;  
 XX KW antiarthritic.  
 XX OS Homo sapiens.  
 XX PN WO200107068-A1.  
 XX PD 01-FEB-2001.  
 XX PF 21-JUL-2000; 2000WO-US020002.  
 XX PR 23-JUL-1999; 99US-0145328P.  
 XX PR 19-JUL-2000; 2000US-00145328.  
 XX PA (UYCA-) UNIV CASE WESTERN RESERVE.  
 XX PI Warman ML;  
 XX DR WPI; 2001-182721/18.  
 XX PT New composition comprising the camptodactyly-arthropathy-coxa vara-pericarditis protein in combination with an anesthetic, useful for treating osteoarthritis, or as lubricants of tissue and joints.  
 XX PS Example 1; Page; 34pp; English.  
 CC The invention relates to a method of treating osteoarthritis via the administration of a composition comprising the camptodactyly-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anaesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder camptodactyly-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the

CC treatment of osteoarthritis. The composition may be applied to reduce the  
CC symptoms of osteoarthritis (e.g., joint pain, loss of range of movement  
CC or joint damage). The present sequence represents human megakaryocyte  
CC stimulating factor (MSF, CAC protein). Note: This sequence is not given  
CC in its entirety in figure 4 of the specification, although a GenBank  
CC accession number was given. This sequence was therefore obtained from  
CC GenBank (U70316)

XX SQ Sequence 1404 AA;

Query Match 99.4%; Score 5339.6; DB 4; Length 1404;  
Best Local Similarity 88.2%; Pred. No. 2.3e-141;  
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLSVVFVIOQVSSQ----- 25  
DB 1 MAWKTLPIYLLLSVVFVIOQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60  
QY 26 -----ELCKGRCFESFERGRCDCDAOCKKYDKCCPDYESFCAB----- 65  
DB 61 KRVTAEALCKGRCFESFERGRCDCDAOCKKYDKCCPDYESFCAEVHNPTSPSPKKAP 120  
QY 66 ----- 65  
DB 121 PPSGASQTIKSTTRSPKPNKKTKVIBSEBITEEHSVSENQESSSSSSSSSSTI 180  
QY 66 -----VKONKQRTKKKTPKPPVVDGAGSLDNGDFKVTTPDTST 106  
DB 181 KIKSSKNSAANRELQKLVKDNKQRTKKKTPKPPVVDGAGSLDNGDFKVTTPDTST 240  
QY 107 TOHNKYSTSPKITTAKPINRPSLPNSDTSKETSIVNKETVETKETTNNKQSTDG 166  
DB 241 TOHNKYSTSPKITTAKPINRPSLPNSDTSKETSIVNKETVETKETTNNKQSTDG 300  
QY 167 KEKTSKAKETQSIKTSKADLAPTSKVLAKEPTTKGPAETTTKGPALTTKPEPTTTPKEPAS 226  
DB 301 KEKTSKAKETQSIKTSKADLAPTSKVLAKEPTTKGPAETTTKGPALTTKPEPTTTPKEPAS 360  
QY 227 TTPKEPTTTIKGAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKEP 286  
DB 361 TTPKEPTTTIKGAPTTKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKEP 420  
QY 287 APPTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 346  
DB 421 APPTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 480  
QY 347 EPAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKPEPAP 406  
DB 481 EPAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKPEPAP 540  
QY 407 TTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT 466  
DB 541 TTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT 600  
QY 467 APAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT 526  
DB 601 APAPTAPKPEPAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT 660  
QY 527 PEEPAPTTPKAAAPNTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT 586  
DB 661 PEEPAPTTPKAAAPNTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT 720  
QY 587 APPTPKPAPKELAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 646  
DB 721 APPTPKPAPKELAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTT 780  
QY 647 TAPTTLKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT 706  
DB 781 TAPTTLKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTT 840  
QY 707 KPAPTTTETPTTTPPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEPVPT 766  
DB 841 KPAPTTTETPTTTPPTTSEVSTPTTTTKEPTTIHKSPDESTPELSAETTPKALENSPKPEPVPT 900

QY 767 TKTBAATKPEMTTAKDKITRDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTTOV 826  
DB 901 TKTBAATKPEMTTAKDKITRDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTTOV 960  
QY 827 TSTTTQDTPFPFKITTLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDORATNSKATTPK 886  
DB 961 TSTTTQDTPFPFKITTLTKTTLAPKVTTKKTTTTEIMNKPEETAKPKDORATNSKATTPK 1020  
QY 887 POKPTKAPKPTSTKPKTTPRVKPTTTPRKMSTTMPBLNPTSRISAEAMLQTTTRN 946  
DB 1021 POKPTKAPKPTSTKPKTTPRVKPTTTPRKMSTTMPBLNPTSRISAEAMLQTTTRN 1080  
QY 947 QTPNSKLVVNPVKSDEAGAGETPHMLLRPHVFPVEVTPDMDYLRVNPQGIINPMLS 1006  
DB 1081 QTPNSKLVVNPVKSDEAGAGETPHMLLRPHVFPVEVTPDMDYLRVNPQGIINPMLS 1140

RESULT 8

ADM98014  
ID ADM98014 standard; protein; 1404 AA.  
XX  
AC ADM98014;  
XX  
DT 01-JUL-2004 (first entry)  
XX  
DE Human megakaryocyte stimulating factor (MSF).

lubricating polypeptide; O-linked oligosaccharide; joint lubrication;  
KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;  
KW megakaryocyte stimulating factor; MSF.

OS Homo sapiens.

XX US2004072741-A1.

XX 15-APR-2004.

XX 02-JUL-2001; 2001US-00897188.

XX 23-APR-1999; 99US-00298970.

XX 24-APR-2000; 2000US-00556246.

XX (JAYG/) JAY G D.

XX Jay GD;

XX WPI; 2004-373948/35.

XX N-PSDB; ADM98015.

XX New tribonectin polypeptides and polynucleotides for lubricating joints  
PT or other tissues to prevent or treat camptodactyl-arthropathy-  
PT pericarditis syndrome or osteoarthritis.

XX Claim 1; SEQ ID NO 1; 34pp; English.

XX The invention relates to a lubricating polypeptide and at least one O-  
CC linked oligosaccharide. The composition and methods are useful for  
CC lubricating joints or other tissues to prevent or treat camptodactyl-  
CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present  
CC sequence represents the amino acid sequence of the human megakaryocyte  
CC stimulating factor (MSF).

XX Sequence 1404 AA;

Query Match 99.4%; Score 5339.6; DB 8; Length 1404;  
Best Local Similarity 88.2%; Pred. No. 2.3e-141;  
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLSVVFVIOQVSSQ----- 25  
DB 1 MAWKTLPIYLLLSVVFVIOQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60

QY 26 -----ELSKGRGCFESPERGREGCDCAQCKYDKCCPDYESFCAE----- 65  
 Db 61 KRVCITAEISCKGRGCFESPERGREGCDCAQCKYDKCCPDYESFCAE VHNPTSPSSKKAP 120  
 QY 66 ----- 65  
 Db 121 PPSGASQTIKSTTKRSPKPNKKTKKVI EEEITEHVSSENQESSSSSSSSSTIW 180  
 QY 66 -----VKDNKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 106  
 Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 240  
 QY 107 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETS LTVNKETIVETKETTITNKQSTDG 166  
 Db 241 TQHNKYSTSPKITTAKPINRPSLPNSDTSKETS LTVNKETIVETKETTITNKQSTDG 300  
 QY 167 KEKTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 226  
 Db 301 KEKTSKAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTPTTPKEPAS 360  
 QY 227 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTPKEPAPTTTKP 286  
 Db 361 TTPKEPTPTTIKSAPTTTPKEPAPTTTKSAPTTTPKEPAPTTTKPAPTTTPKEPAPTTTKP 420  
 QY 287 APTTTKSAPTTTPKEPAPTTPKAPPTTPKEPAPTTTPKEPTPTTPKEPAPTTTPKEPAPTTPK 346  
 Db 421 APTTTKSAPTTTPKEPAPTTPKAPPTTPKEPAPTTTPKEPTPTTPKEPAPTTTPKEPAPTTPK 480  
 QY 347 EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTTTPKEPAP 406  
 Db 481 EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTTTPKEPAP 540  
 QY 407 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKPAPTTTKP 466  
 Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKPAPTTTKP 600  
 QY 467 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 526  
 Db 601 APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 560  
 QY 527 PEEAPPTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 586  
 Db 661 PEEAPPTTPKAAAPNTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP 720  
 QY 587 APTTPKAPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 646  
 Db 721 APTTPKAPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 780  
 QY 647 TAPTTLKEPAPTTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 706  
 Db 781 TAPTTLKEPAPTTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 840  
 QY 707 KPAPTTTPETPTTSEVSPPTTKPTTHKSPDSTPESLAPPTPKALENPKGPGVPT 766  
 Db 841 KPAPTTTPETPTTSEVSPPTTKPTTHKSPDSTPESLAPPTPKALENPKGPGVPT 900  
 QY 767 TKTPAATKPEMTTAKDKTTERDLRTPTTTTAAAPKMTKETAATTTKTTESKITATTQV 826  
 Db 901 TKTPAATKPEMTTAKDKTTERDLRTPTTTTAAAPKMTKETAATTTKTTESKITATTQV 960  
 QY 827 TSTTTQDTPPKITLTKTTTLAPKVTTKTTITTEIMNKPEETAKPKDRATNSKATTPK 886  
 Db 961 TSTTTQDTPPKITLTKTTTLAPKVTTKTTITTEIMNKPEETAKPKDRATNSKATTPK 1020  
 QY 887 POKPTKAPKPTSTTKPKTMPVRPKPTTPTRKMTSNMPELNPTSRIAEAMLQTTTPN 946  
 Db 1021 POKPTKAPKPTSTTKPKTMPVRPKPTTPTRKMTSNMPELNPTSRIAEAMLQTTTPN 1080  
 QY 947 QTPNSKLVNPKSDAGGAEGETPHMLLRPHVFMPEVPTPDMDYLPVFNQGIINPMLS 1006  
 Db 1081 QTPNSKLVNPKSDAGGAEGETPHMLLRPHVFMPEVPTPDMDYLPVFNQGIINPMLS 1140

RESULT 9  
 AAU32262  
 ID AAU32262 standard; protein; 1415 AA.  
 XX AC AAU32262;  
 XX DT 18-DEC-2001 (first entry)  
 XX DE Novel human secreted protein #2753.  
 XX KW Human; vaccination; gene therapy; nutritional supplement;  
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.  
 XX OS Homo sapiens.  
 XX WO200179449-A2.  
 XX PD 25-OCT-2001.  
 XX PF 16-APR-2001; 2001WO-US0008656.  
 XX PR 18-APR-2000; 2000US-00552929.  
 XX PR 26-JAN-2001; 2001US-00770160.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Drmanac RT;  
 XX WI MPI; 2001-611725/70.  
 XX PT Nucleic acids encoding a range of human polypeptides, useful in genetic  
 XX vaccination, testing and therapy.  
 XX PS Claim 20; Page 573; 765pp; English.  
 XX CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising the  
 CC nucleic acids encoding the polypeptides and cells genetically engineered  
 CC to express them are also useful for producing the proteins. The proteins  
 CC are useful in genetic vaccination, testing and therapy, and can be used  
 CC as nutritional supplements. They may be used to increase stem cell  
 CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon  
 CC and/or nerve tissue growth or regeneration; immune suppression and/or  
 CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.  
 CC AAU29510-AAU33304 represent the amino acid sequences of novel human  
 CC secreted proteins of the invention  
 XX SQ Sequence 1415 AA;

Query Match 99.4%; Score 5339.6; DB 4; Length 1415;  
 Best Local Similarity 88.2%; Pred. No. 2.3e-141;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ----- 25  
 Db 12 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDRATNCNDYNCOHMBCCPDF 71  
 QY 26 -----ELSKGRGCFESPERGREGCDCAQCKYDKCCPDYESFCAE----- 65  
 Db 72 KRVCITAEISCKGRGCFESPERGREGCDCAQCKYDKCCPDYESFCAE VHNPTSPSSKKAP 131  
 QY 66 ----- 65  
 Db 132 PPSGASQTIKSTTKRSPKPNKKTKKVI EEEITEHVSSENQESSSSSSSSSTIW 191  
 QY 66 -----VKDNKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 106

Db	192	KIKSSKNSAANRELOKLVKDNKNRKKTKPPVVDDEAGSLDNGDFKVTTPDST	251
Qy	107	TOHNKVSTSPKLTITAKPINRSLPNSDTSKETSITVKNKETTVEKTTTNKQSTDG	166
Db	252	TOHNKVSTSPKLTITAKPINRSLPNSDTSKETSITVKNKETTVEKTTTNKQSTDG	311
Qy	167	KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKPSTPTTPKEPAS	226
Db	312	KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKPSTPTTPKEPAS	371
Qy	227	TPKPEPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTKP	286
Db	372	TPKPEPTTIKSAPTTPKEPAPTTTKSAPTTTPKEPAPTTTPKEPAPTTTKP	431
Qy	287	APTTPKSAPTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTK	346
Db	432	APTTPKSAPTTPKEPAPTTTPKAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTK	491
Qy	347	EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAP	406
Db	492	EPAPTAPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAP	551
Qy	407	TTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP	466
Db	552	TTKSAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP	611
Qy	467	APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT	526
Db	612	APTAPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT	671
Qy	527	PESAPTTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP	586
Db	672	PESAPTTPKAAAPNTPKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKP	731
Qy	587	APTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP	646
Db	732	APTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP	791
Qy	647	TAPTTLKEPAPTTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP	706
Db	792	TAPTTLKEPAPTTTPKAPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP	851
Qy	707	KPAPTTPETPPPTSEVSTPTTKPTTIHKSPDSTPSPKAPPTTPKALENSPKEPAP	766
Db	852	KPAPTTPETPPPTSEVSTPTTKPTTIHKSPDSTPSPKAPPTTPKALENSPKEPAP	911
Qy	767	TKTPAATKPMWTTAKDKTTERDLRTTPETTTAAPKMTKETAATTEKTTESKITATTTQV	826
Db	912	TKTPAATKPMWTTAKDKTTERDLRTTPETTTAAPKMTKETAATTEKTTESKITATTTQV	971
Qy	827	TSITTTQDTTPFKITLTLATTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK	886
Db	972	TSITTTQDTTPFKITLTLATTTLAPKVTTKKTIITTEIMNKPEETAKPKDRATNSKATTPK	1031
Qy	887	PQKPTKAPKPTSTKPKTPRVRUKPTTTPRKTSTMPKELAPTSIAEAMLOTITPRN	946
Db	1032	PQKPTKAPKPTSTKPKTPRVRUKPTTTPRKTSTMPKELAPTSIAEAMLOTITPRN	1091
Qy	947	QTNKSLVENVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLRVPVNOGIIINPMLS	1006
Db	1092	QTNKSLVENVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLRVPVNOGIIINPMLS	1151
RESULT 10			
ID	ADK65839		
XX	ADK65839	standard; protein; 1404 AA.	
AC	ADK65839;		
XX	AC		
DT	06-MAY-2004	(first entry)	
XX	XX		
DE	Angiogenesis-differentially expressed protein #53.		

XX	cytostatic; cardiant; vasotropic; antiarteriosclerotic;		
KW	angiogenesis inhibitor; angiogenesis stimulator; angiogenic index;		
KW	gene expression; cancer; coronary artery disease; myocardial ischemia;		
KW	coronary arteriosclerosis; forensic medicine.		
OS	Homo sapiens.		
XX	WO2003066831-A2.		
FN	14-AUG-2003.		
PD	07-FEB-2003; 2003WO-US003848.		
XX	07-FEB-2002; 2002US-00067482.		
PF	10-JUN-2002; 2002US-00164595.		
XX	16-AUG-2002; 2002US-0403649P.		
PR	03-JAN-2003; 2003US-0437746P.		
XX	(ORIG-) ORIGENE TECHNOLOGIES INC.		
PA	Sun Z, Li X, Kovacs KF, Fan W, Jay G;		
XX	WPI; 2003-731502/69.		
PI	Determining the angiogenic index of a tissue or cell sample using		
XX	expression levels of differentially expressed genes, useful for		
XX	diagnosing or treating cancer, coronary artery disease, myocardial		
XX	ischemia and/or arteriosclerosis.		
XX	Disclosure; SEQ ID NO 78; 296pp; English.		
PS	The invention relates to a method of determining the angiogenic index of		
CC	a tissue or cell sample comprising assessing, in a sample, the expression		
CC	levels of one or more differentially-expressed gene from any of 34 DNA		
CC	sequences, given in the specification, where the levels are indicative of		
CC	the angiogenic index. The methods and compositions of the present		
CC	invention are useful for diagnosing, preventing and/or treating cancer,		
CC	coronary artery disease, myocardial ischemia or coronary		
CC	arteriosclerosis. They can also be used in research, drug discovery and		
CC	forensic medicine involving angiogenesis. This sequence corresponds to		
CC	one of the differentially expressed proteins of the invention.		
XX	Sequence 1404 AA;		
QY	Query Match 99.2%; Score 5330.6; DB 7; Length 1404;		
DB	Best Local Similarity 88.1%; Pred. No. 4.1e-141;		
DB	Matches 1004; Conservative 0; Mismatches 2; Indels 134; Gaps 2;		
QY	1 MANKTLPYVLLLLSVFVIOQVSSQ-----	25	
DB	1 MANKTLPYVLLLLSVFVIOQVSSQDLSACGRCGEGYSRATCNCNDYHCQHYMECCPDF	60	
QY	26 -----ELSCGRCFCFSFERGRCDCDAQCKYDKKCCPDYESPCAE-----	65	
DB	61 KRVCTAELSCGRCFCFSFERGRCDCDAQCKYDKKCCPDYESFCAEVHNFTSPSSKAP	120	
QY	66 -----	65	
DB	121 PPSGASQTIKSTTKRSPKPNKKTKVIESEITEHVSVENQSSSSSSSSSSSSSTIR	180	
QY	66 -----VKDNKNRRTKKKPKPPVVDDEAGSLDNGDFKVTTPDST	106	
DB	181 KIKSSKNSAANRELOKLVKDNKNRRTKKKPKPPVVDDEAGSLDNGDFKVTTPDST	240	
QY	107 TOHNKVSTSPKLTITAKPINRSLPNSDTSKETSITVKNKETTVEKTTTNKQSTDG	166	
DB	241 TOHNKVSTSPKLTITAKPINRSLPNSDTSKETSITVKNKETTVEKTTTNKQSTDG	300	
QY	167 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKPSTPTTPKEPAS	226	
DB	301 KEKTTSAKETQSIKTSKDLAPTSKVLAKPTPKAETTTKGPALTTPKPSTPTTPKEPAS	360	





```

ABUS3253
ID ABUS3253 standard; protein; 551 AA.
XX
AC ABUS3253;
XX
DT 14-APR-2003 (first entry)
XX
DE Human testes-derived DKFphtes3_4019 homologue #2.
XX
KW Human; gene therapy; vaccine; disease treatment; detection.
XX
OS Homo sapiens.
XX
PN WO200112659-A2.
XX
PD 22-FEB-2001.
XX
PF 18-AUG-2000; 2000WO-IB001496.
XX
PR 18-AUG-1999; 99US-0149499P.
PR 28-SEP-1999; 99US-0156503P.
XX
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI Wiemann S;
XX
DR WPI; 2001-327840/34.
XX
PT Nucleic acids having the sequences of clones isolated from libraries of
FT different human tissues, useful in recombinant DNA methodologies.
XX
PS Example III; Page 892-893; 1095pp; English.
XX
CC This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 551 AA;

Query Match 53.8%; Score 2886; DB 4; Length 551;
Best Local Similarity 100.0%; Pred. No. 2.7e-73;
Matches 544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 463 TKKPATAPKEPAPTPKETAPTPKXLTPTTPEKLAPTPKEPAPTPPELAPTTPEP 522
DB 1 TKKPATAPKEPAPTPKETAPTPKXLTPTTPEKLAPTPKEPAPTPPELAPTTPEP 60

QY 523 TPTTPEPAPTPPKAAAPNTPKEPAPTPPKETAPTPKEPAPTPPKETAPTPKGTAPTT 582
DB 61 TPTTPEPAPTPPKAAAPNTPKEPAPTPPKETAPTPKEPAPTPPKETAPTPKGTAPTT 120

QY 583 LKEPAPTPKKAPKELAPTTTKEPTSTTSDDKAPTPPKGTADTPPKAPAPTPPKAPPT 642
DB 121 LKEPAPTPKKAPKELAPTTTKEPTSTTSDDKAPTPPKGTADTPPKAPAPTPPKAPPT 180

QY 643 TPKGATPTLLKEPAPTPPKKAPKELAPTTTKEPTSTTSDDKAPTPPKGTADTPPKAPAP 702
DB 181 TPKGATPTLLKEPAPTPPKKAPKELAPTTTKEPTSTTSDDKAPTPPKGTADTPPKAPAP 240

QY 703 TTPKKAPAPTPPPPTTSVSTPTTKEPTTHKSPDESTPBLSPAPTPPKALENSPKPEP 762
DB 241 TTPKKAPAPTPPPPTTSVSTPTTKEPTTHKSPDESTPBLSPAPTPPKALENSPKPEP 300

QY 763 GVPVTTKTPAAKPEMTTAKDKTTERDLRTPPTTAAAPKMTKETATTTTETTESKITAT 822
DB 301 GVPVTTKTPAAKPEMTTAKDKTTERDLRTPPTTAAAPKMTKETATTTTETTESKITAT 360

```

```

QY 823 TTQVSTTTQDTPPKITTLTKTTLAPKVTITTKITITTTIMNKPETAKPKDRATNSKA 882
DB 361 TTQVSTTTQDTPPKITTLTKTTLAPKVTITTKITITTTIMNKPETAKPKDRATNSKA 420
QY 883 TTPKPKQPTKAPKKPTSTTKKPKTMPRVKPKTTPTPRKVMTSTMPELNPTSRIAEAMLOTT 942
DB 421 TTPKPKQPTKAPKKPTSTTKKPKTMPRVKPKTTPTPRKVMTSTMPELNPTSRIAEAMLOTT 480
QY 943 TRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIIN 1002
DB 481 TRPNQTPNSKLVEVNPKSEDAAGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIIN 540
QY 1003 PMLS 1006
DB 541 PMLS 544

RESULT 13
AAB29778
ID AAB29778 standard; protein; 902 AA.
XX
AC AAB29778;
XX
DT 28-FEB-2001 (first entry)
XX
DE Human MSF-derived tribonection.
XX
KW Human tribonection; MSF; megakaryocyte stimulating factor;
KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;
KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;
KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.
XX
OS Homo sapiens.
XX
PN WO200064930-A2.
XX
PD 02-NOV-2000.
XX
PF 24-APR-2000; 2000WO-US010953.
XX
PR 23-APR-1999; 99US-00298970.
XX
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
PI Jay GD;
XX
DR WPI; 2001-024673/03.
XX
PT Novel tribonection polypeptide useful as lubricant for treating
PT osteoarthritis, comprises O-linked lubricating moiety.
XX
PS Disclosure; Fig 1; 47pp; English.
XX
CC The invention relates to a human tribonection which is a product of
CC alternative splicing of the human MSF (megakaryocyte stimulating factor)
CC gene. The tribonection has at least one O-linked oligosaccharide
CC lubricating moiety and has a polypeptide sequence comprising 1-76 repeats
CC of a motif having at least 50% identity to the sequence KEPAPPT
CC (AAB29774). The invention also relates to a nucleic acid encoding a human
CC MSF-derived tribonection; a biocompatible composition comprising a human
CC tribonection for inhibiting tissue adhesion formation; and a method of
CC diagnosing osteoarthritis or a predisposition to osteoarthritis by
CC measuring the amount of MSF or its fragment in a biological sample of a
CC mammal, wherein an increased amount of MSF compared to a control
CC indicates the presence of or predisposition to developing osteoarthritis.
CC The tribonection and DNA encoding it are useful in the treatment of
CC osteoarthritis, where they may be used for lubricating mammalian joints,
CC such as articulating joints of humans, dogs or horses. The tribonection,
CC when formulated as a membrane, foam, gel or fibre, is useful for
CC inhibiting adhesion between two surfaces such as the injured tissues of a
CC mammal, where the injury is caused by a surgical insertion or trauma, or
CC an artificial device e.g., an orthopaedic implant. In particular, one of

```

CC the surfaces is pericardial tissue. DNA encoding a tribonectin may be  
CC used in gene therapy. The present sequence represents a substantial  
CC portion of a human MSF-derived tribonectin  
XX  
SQ Sequence 902 AA;  
Query Match 53.6%; Score 2880.1; DB 4; Length 902;  
Best Local Similarity 71.4%; Pred. No. 7.9e-73;  
Matches 675; Conservative 27; Mismatches 74; Indels 169; Gaps 66;  
QY 66 VKONKKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNVKVSPTKITTAKPIN 125  
Db 1 VKONKKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNVKVSPTKITTAKPIN 60  
QY 126 PRSLPNSDTSKETSITVWNETTETKTTTNNKQTSIDGKEKTTSAKETQSIKTSK 185  
Db 61 PRSLPNSDTSKETSITVWNETTETKTTTNNKQTSIDGKEKITSKETQSIKTSK 120  
QY 186 DLAPTSKVLAKPTPKAETTTKGPALTTPEKPTPTTPEKPASTTPKGPPTTTIKSAPTTPK 245  
Db 121 DLAPTSKVLAKPTPKAETTTKGPAL-----TTPEKPASTTPKGPPTTTIKSAPTTPK 172  
QY 246 EPAPTTTKSAPTTKGPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 305  
Db 173 EPAPTTTKSAPTTKGPAP--TTKEPAPTT--KEPAP--TTKEPAPTTTKEPAPTT--KEPAPTT 228  
QY 306 PKKPAPTTKEPAPTTPKPTPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 365  
Db 229 -KEPAPTT--KEPAPTT--KEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT--KEP 281  
QY 366 PTTPEKAPTTTKEPSPTTPKBPATTTKSAPTTTTKEPAPTTTKEPAPTTTKEPSPTTTKE 425  
Db 282 PTT--KEPAP--TTKEPAPTT--KEPAPTTKBPAP--TTKEPAPTTTKEPAPTT--KEPAP--TTKE 335  
QY 426 PAPTTPKEPAPTTKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 485  
Db 336 PAPT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAP--TTKEPAPTT--KEPAPTT--KEPAPT 388  
QY 486 TPKKLTPTPEKAPTTPEKAPTTPEELAPTTPEPTTPEPAPTTTPEPAPTTTPEPAPTTPE 545  
Db 389 T--KEPAPTTTKEP--APTKE--PAPT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPT--KE 440  
QY 546 PAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 603  
Db 441 PAPT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTTTKEPAPTT--KEPAPT 493  
QY 604 TKEPTSTSDKAPTTTKEGAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 663  
Db 494 TKEPAPTTTKE--PAPT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPTT--KEPAPT--KEP 546  
QY 664 AP--KEAPTTTKEGPTSTSDKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 721  
Db 547 APTTKEPAP--TTKEPAPTTTKE--PAPT--KEPAPTT--KEPAPTT--KEPAPTTTKE--PAPT 600  
QY 722 EVSTPTTKEPTTIHKSPOESTPELSAETPKALENSPKPEGVPTTKTIPAAKPEMTTUA 781  
Db 601 EPA--PTTKEPAPTTKEPAPT--KEPAP-----TTKEP--APTKEPAPT----- 640  
QY 782 KOKTTERDLRTPETTTAAAPKMTKEPATTTTEKTTESKITATTTQVTSITTOQTTPPKITT 841  
Db 641 -----TTKEPAPT-----KEPAPT-----KEPAPT----- 667  
QY 842 LKTTTLAPKVTTTKTITTEIMNKBEETAKPKDRATNSKAITPKQKQTKAPKKTSTK 901  
Db 668 -----TTKEPAPT-----TTKEPAPT----- 684  
QY 902 KPKTMRVRKPKTTPTRKWTSTMPBLNPTSRIAEAMLQTTTRPNQTSKLVNPKSE 961  
Db 685 EP-----APTTPTRKWTSTMPBLNPTSRIAEAML--TTTRPNQTSKLVNPKSE 735  
QY 962 DAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1006  
Db 736 DAGGAGETPHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 780

RESULT 14  
ABU53254  
ID ABU53254 standard; protein; 513 AA.  
XX  
AC ABU53254;  
XX  
DT 14-APR-2003 (first entry)  
XX  
DE Human testes-derived DKFZphtes3\_4019 homologue #3.  
XX  
KW Human; gene therapy; vaccine; disease treatment; detection.  
XX  
OS Homo sapiens.  
XX  
PN WO200112659-A2.  
XX  
PD 22-FEB-2001.  
XX  
PF 18-AUG-2000; 2000WO-1B001496.  
XX  
PR 18-AUG-1999; 99US-0149499P.  
PR 28-SEP-1999; 99US-0156503P.  
XX  
PA (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX  
PI Wiemann S;  
XX  
DR WPI; 2001-327840/34.  
XX  
PT Nucleic acids having the sequences of clones isolated from libraries of  
PT different human tissues, useful in recombinant DNA methodologies.  
XX  
PS Example III; Page 893; 1095pp; English.  
XX  
CC This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence represents a homologue  
CC of a polypeptide described in the disclosure of the invention  
XX  
SQ Sequence 513 AA;

Query Match 51.3%; Score 2757; DB 4; Length 513;  
Best Local Similarity 100.0%; Pred. No. 1.1e-69;  
Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 73 RTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNVKVSPTKITTAKPINPRSLPP 132  
Db 1 RTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDSTTQHNVKVSPTKITTAKPINPRSLPP 60  
QY 133 NSDTSKETSITVWNETTETKTTTNNKQTSIDGKEKTTSAKETQSIKTSKADLAPT 192  
Db 61 NSDTSKETSITVWNETTETKTTTNNKQTSIDGKEKTTSAKETQSIKTSKADLAPT 120  
QY 193 VLAKEPTKAEITTKGPALTTPEKPTPTTPEKPASTTPKGPPTTTIKSAPTTPKEPAPT 252  
Db 121 VLAKEPTKAEITTKGPALTTPEKPTPTTPEKPASTTPKGPPTTTIKSAPTTPKEPAPT 180  
QY 253 KSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 312  
Db 181 KSAPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPT 240  
QY 313 TPKEPAPTTTKEPPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 372  
Db 241 TPKEPAPTTTKEPPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 300

QY 373 APTTTKEPSPPTPKPAPTTTKSAPTTTKPAPTTTKSAPTTTKPAPTTTKPAPTTTK 432  
 Db |||||  
 QY 301 APTTTKEPSPPTPKPAPTTTKSAPTTTKPAPTTTKSAPTTTKPAPTTTKPAPTTTK 360  
 Db |||||  
 QY 433 EPAPTTPKPAPTTPKPAPTTPKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTK 492  
 Db |||||  
 QY 361 EPAPTTPKPAPTTPKPAPTTPKPAPTTTKPAPTTTKPAPTTTKPAPTTTKPAPTTTK 420  
 Db |||||  
 QY 493 TTPKELAPTTPKPAPTTPEELAPTTPEPTPTTPEPAPTTPKAAAPNTPKPAPTTTK 552  
 Db |||||  
 QY 421 TTPKELAPTTPKPAPTTPEELAPTTPEPTPTTPEPAPTTPKAAAPNTPKPAPTTTK 480  
 Db |||||  
 QY 553 EPAPTTPKPAPTTPKPAPTTPKGTAPTTLKE 585  
 Db |||||  
 QY 481 EPAPTTPKPAPTTPKPAPTTPKGTAPTTLKE 513  
 Db |||||  
 RESULT 15  
 AAR80041  
 ID AAR80041 standard; protein; 452 AA.  
 XX  
 AC AAR80041;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 10-APR-1996 (first entry)  
 XX  
 DE Human megakaryocytopoietin protein.  
 XX  
 KW Human; megakaryocytopoietin; wheat germ agglutinin; heparin;  
 KW megakaryocyte; aplastic anaemia; thrombocytopenia; hematocytopenia;  
 KW multipotential stem cell.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 393..396 /note= "unspecified amino acids"  
 FT Misc-difference 444..446 /note= "unspecified amino acids"  
 FT  
 FT  
 XX W09523861-AL.  
 XX  
 PD 08-SEP-1995.  
 XX  
 PF 06-MAR-1995; 95WO-CN000015.  
 XX  
 PR 04-MAR-1994; 94CN-00112066.  
 XX  
 PA (SHAN-) SHANGHAI BEITE BIOTECHNOLOGY CO LTD.  
 XX  
 PI Gu X, Han Z, Shen Q;  
 XX  
 DR WPI; 1995-320576/41.  
 DR N-PSDB; AAT04546.  
 XX  
 PT New haematopoietic cell growth factor - used for treating  
 PT thrombocytopenia and hematocytopenia.  
 XX  
 PS Example; Page 23; 36pp; Chinese.  
 XX

CC This sequence represents the human megakaryocytopoietin (MPO) protein.  
 CC This sequence was purified using a carrier which can couple wheat germ  
 CC agglutinin and heparin to separate MPO. Fragments of this sequence (see  
 CC AAR80039 and AAR80040) were used to produce the amplification primers  
 CC shown in AAT04544 and AAT04545. The fragments amplified by these primers  
 CC can then be used as probes to screen human cDNA libraries for MPO cDNA.  
 CC The MPO cDNA can then be inserted into a plasmid which is used to  
 CC transform cells to produce MPO. The MPO sequence is capable of promoting  
 CC colony formation of megakaryocytes, enlarging the size of megakaryocytes  
 CC and stimulating the proliferation of multipotential stem cells. The  
 CC factor may be used for treating thrombocytopenia and hematocytopenia. The  
 CC purification method can be used to isolate MPO from human urine or serums  
 CC of patients with aplastic anaemia, and from animal blood or urine by

CC radiation exposing the animals to induce aplastic anaemia. (Updated on 25  
 CC -MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 452 AA;

Query Match 31.5%; Score 1694.8; DB 2; Length 452;  
 Best Local Similarity 72.2%; Pred. No. 4.8e-40;  
 Matches 328; Conservative 3; Mismatches 21; Indels 102; Gaps 3;  
 QY 1 MAWKTLPIYLLLLLSVFVIQQVSSQ----- 25  
 Db 1 MAWKTLPIYLLLLLSVFVIQQVSSQDLSSCAGCGEGYGRDATCNCYNCQHYMECCPDF 60  
 QY 26 -----ELSCGRCFESFERGECDCDAOCKYDKCCPDYESFCA----- 64  
 Db 61 KRVCTAELSCGRCFESFERGECDCDAOCKYDKCCPDYESFCAEVHNFTSPSSKKAP 120  
 QY 65 -----FVKONKKNRTKKKTPKPPVVVDEAG 89  
 Db 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIESEIEVKNKKQRTKKKTPKPPVVVDEAG 180  
 QY 90 SGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSPSLPNSDTSKETSILVNKETT 149  
 Db 181 SGLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSPSLPNSDTSKETSILVNKETT 240  
 QY 150 VETKETTTTNKOTSTDGKEKTTSAKETQSIKETSADLAPTSKVLAKPTPKAETTTKGPA 209  
 Db 241 VETKETTTTNKOTSTDGKEKTTSAKETQSIKETSADLAPTSKVLAKPTPKAETTTKGPA 300  
 QY 210 LTTPKPPTTTPKEPASTTPKBPPTTIKSAPTTPKPAPTTTKSAPTTPKPAPTTTKE 269  
 Db 301 LTTPKPPTTTPKEPASTTPKBPPTTIKSAPTTPKPAPTTTKSAPTTPKPAPTTTKE 360  
 QY 270 PAPTTPKEPAPTTTKPAPTTTKSAPTTPKPAPTTTKPAPTTTKPAPTTTKEPTT 329  
 Db 361 PAPTTPKEPAPTTTKPAPTTTKSHPLPRSCXXXTQ-----TPKEHPPL 409  
 QY 330 PKEPAPTTKEPAPTTTKPAPTTAPKKPAPTTPKP 363  
 Db 410 PRSLHPTKEPAPTTTKPAPTTAPKKPAPLPLE 443

Search completed: October 13, 2004, 11:37:08  
 Job time : 90.1065 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 17.4873 Seconds  
(without alignments)  
3815.116 Million cell updates/sec

Title: SEQ1-E  
Perfect score: 5373  
Sequence: 1 MAWKTLPIYLLLSVFIQ.....DMDYLPRVFNQGIINPMLS 1006

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/6A COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/6B COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/PCUS COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfilesi.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5373	100.0	1038	4	US-07-757-022B-74
2	5373	100.0	1270	4	US-07-757-022B-44
3	5358.9	99.7	1311	4	US-07-757-022B-42
4	5358.7	99.7	1049	4	US-07-757-022B-58
5	5358.7	99.7	1313	4	US-07-757-022B-142
6	5358	99.7	1320	4	US-07-757-022B-46
7	5358	99.7	1320	4	US-07-757-022B-60
8	5353.7	99.6	1363	4	US-07-757-022B-52
9	5349	99.6	1320	4	US-10-164-595-58
10	5344.6	99.5	1354	4	US-07-757-022B-48
11	5343.9	99.5	1361	4	US-07-757-022B-40
12	5339.6	99.4	1140	4	US-07-757-022B-104
13	5339.6	99.4	1404	4	US-07-757-022B-2
14	5339.6	99.4	1404	4	US-07-757-022B-62
15	5338.6	99.4	1404	4	US-09-298-970A-1
16	5330.6	99.2	1404	4	US-10-164-595-78
17	5241.8	97.6	1022	4	US-07-757-022B-84
18	5232.6	97.4	1314	4	US-07-757-022B-50
19	5011	93.3	941	4	US-07-757-022B-14
20	1335.3	24.9	5179	4	US-09-538-092-1258
21	1075.7	20.0	8991	4	US-08-714-741-32
22	803.4	15.0	3256	4	US-09-919-172-98
23	803.4	15.0	3256	4	US-09-976-594-22
24	803.4	15.0	3256	4	US-09-919-039-21
25	795	14.8	3118	3	US-09-578-181-1
26	790.9	14.7	2972	3	US-09-579-181-2
27	757.4	14.1	1837	3	US-08-928-361B-5

Sequence 5, Appli  
Sequence 1142, Ap  
Sequence 6, Appli  
Sequence 6, Appli  
Sequence 5, Appli  
Sequence 4, Appli  
Sequence 30227, A  
Sequence 4, Appli  
Sequence 3, Appli  
Sequence 17307, A  
Sequence 425, App  
Sequence 3, Appli  
Sequence 726, App  
Sequence 1135, Ap  
Sequence 1262, Ap  
Sequence 5, Appli  
Sequence 98, Appl  
Sequence 110, App

## ALIGNMENTS

RESULT 1  
US-07-757-022B-74  
; Sequence 74, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/07757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 08-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 74:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1038 amino acids  
; TYPE: AMINO ACID

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-74

Query Match 100.0%; Score 5373; DB 4; Length 1038;  
Best Local Similarity 100.0%; Pred. No. 1.5e-167; Indels 0; Gaps 0;  
Matches 100%; Conservative 0; Mismatches 0;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60

QY 61 SFCAEVKDNKNKRTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHNKYSTSPKITT 120  
DB 61 SFCAEVKDNKNKRTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHNKYSTSPKITT 120

QY 121 AKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDGKEKTTSAKETQSIE 180  
DB 121 AKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDGKEKTTSAKETQSIE 180

QY 181 KTSADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKBPASTTTPKEPTTTIKSA 240  
DB 181 KTSADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKBPASTTTPKEPTTTIKSA 240

QY 241 PTPKBPATTTKSAPTTKBPATTTKBPATTKBPATTKBPATTKBPATTKBPATTKBPATTK 300  
DB 241 PTPKBPATTTKSAPTTKBPATTTKBPATTKBPATTKBPATTKBPATTKBPATTKBPATTK 300

QY 301 PAPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 360  
DB 301 PAPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 360

QY 361 PKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 420  
DB 361 PKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 420

QY 421 TTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 480  
DB 421 TTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 480

QY 481 ETAPTTPKLTPTTPEKLAFTTPEKAPTTTPEELAPTTPPEELAPTTPPEELAPTTPPEELAPTTP 540  
DB 481 ETAPTTPKLTPTTPEKLAFTTPEKAPTTTPEELAPTTPPEELAPTTPPEELAPTTPPEELAPTTP 540

QY 541 NTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 600  
DB 541 NTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 600

QY 601 PTTTKEPTSTSDKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 660  
DB 601 PTTTKEPTSTSDKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 660

QY 661 KKPAPKELAPTTTKGPTSTSDKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 720  
DB 661 KKPAPKELAPTTTKGPTSTSDKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTT 720

QY 721 SEVSTPTTTPKEPTTIHNSPDESTPELSABPTPKALENSPKBPAGVPTTKTAAKPEMTTIT 780  
DB 721 SEVSTPTTTPKEPTTIHNSPDESTPELSABPTPKALENSPKBPAGVPTTKTAAKPEMTTIT 780

QY 781 AKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQVTSITTTQDTTPPKIT 840  
DB 781 AKDKTTERDLRTTPETTTAAPKMTKETATTTTEKTTESKITATTTQVTSITTTQDTTPPKIT 840

QY 841 TLKTTTILAPKVTTTKITTTTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKPTST 900  
DB 841 TLKTTTILAPKVTTTKITTTTEIMNKPEETAAPKDRATNSKATTPKQKPTKAPKPTST 900

QY 901 KKPATMPRVKPKTTPTPRKMSTNMPNLNPTSRABEAMLOTTTPRPNQTPNSKLVEVNPKS 960  
DB 901 KKPATMPRVKPKTTPTPRKMSTNMPNLNPTSRABEAMLOTTTPRPNQTPNSKLVEVNPKS 960

QY 961 EDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1006

DB 961 EDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1006

RESULT 2

US-07-757-022B-44  
Sequence 44, Application US/07757022B  
Patent No. 6433142  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELEPHONE: (617) 876-1170  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1270 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-44

Query Match 100.0%; Score 5373; DB 4; Length 1270;  
Best Local Similarity 100.0%; Pred. No. 2e-167; Indels 0; Gaps 0;  
Matches 100%; Conservative 0; Mismatches 0;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60  
DB 1 MAWKTLPIYLLLLSVFVIQVSSQELSCKGRCFESFERGRCDCDAQCKKYDKCCPDYE 60

QY 61 SFCAEVKDNKNKRTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHNKYSTSPKITT 120  
DB 61 SFCAEVKDNKNKRTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHNKYSTSPKITT 120

QY 121 AKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDGKEKTTSAKETQSIE 180  
DB 121 AKPINRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDGKEKTTSAKETQSIE 180

```
QY 181 KTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPASTTTPKEPTTTIKSA 240
Db 181 KTSADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKBPASTTTPKEPTTTIKSA 240
QY 241 PTTKPEAPTTTSAPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 300
Db 241 PTTKPEAPTTTSAPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 300
QY 301 PATTTPKAPPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 360
Db 301 PATTTPKAPPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 360
QY 361 PKEPAPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 420
Db 361 PKEPAPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 420
QY 421 TTTKPEAPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 480
Db 421 TTTKPEAPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 480
QY 481 ETAPTTPKKLTPTTPEKLAETTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 540
Db 481 ETAPTTPKKLTPTTPEKLAETTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 540
QY 541 NTTPKEAPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 600
Db 541 NTTPKEAPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 600
QY 601 PTTTKEPTSTTSKAPPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 660
Db 601 PTTTKEPTSTTSKAPPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 660
QY 661 KKPAPKELAPTTTGGTSTTSKAPPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 720
Db 661 KKPAPKELAPTTTGGTSTTSKAPPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 720
QY 721 SEVSTPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 780
Db 721 SEVSTPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 780
QY 781 AKDKTERDLRTTTPETTTAAPKMTKETATTTKTTESKITATTTQVSTTTTQDTTTPFKIT 840
Db 781 AKDKTERDLRTTTPETTTAAPKMTKETATTTKTTESKITATTTQVSTTTTQDTTTPFKIT 840
QY 841 TLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKORATNSKATTPKPKQPTTAPKPKPTST 900
Db 841 TLKTTTLAPKVTTKKTIITTEIMNKPEETAKPKORATNSKATTPKPKQPTTAPKPKPTST 900
QY 901 KKPKTMPVRVKPTTTPPKMTSTWPELNPTSRPAEAMLOTTTRNPTNSKLVNPKS 960
Db 901 KKPKTMPVRVKPTTTPPKMTSTWPELNPTSRPAEAMLOTTTRNPTNSKLVNPKS 960
QY 961 EDAGAGETPHMLRPHVMEVPTPDMDLPRVNOGIIINPMLS 1006
Db 961 EDAGAGETPHMLRPHVMEVPTPDMDLPRVNOGIIINPMLS 1006

RESULT 3
US-07-757-022B-42
; Sequence 42, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
```

```
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Pacencin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Juann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-42
```

```
Query Match 99.7%; Score 5358.9; DB 4; Length 1311;
Best Local Similarity 96.1%; Pred. No. 5.8e-167;
Matches 1006; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKLPYILLLLSVFVIQVSSQ-----25
Db 1 MAWKLPYILLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNDYNCQHYMECCPDF 60
QY 26 -----ELSCGRCFESFERGECDCDAOCKKYDKCCPDYESFCAEVKDNKQRTKKKPT 79
Db 61 KKVCTAEISCKGRCFESFERGECDCDAOCKKYDKCCPDYESFCAEVKDNKQRTKKKPT 120
QY 80 KPFPVVDGAGSLDNGDFKVTTPDTSTTQHNVKSTSPKITTAKPINRPSLPNPSDTSKE 139
Db 121 KPFPVVDGAGSLDNGDFKVTTPDTSTTQHNVKSTSPKITTAKPINRPSLPNPSDTSKE 180
QY 140 TSLTVNKETTIVETKETTNTNKQTSDDGKEKTSKAKTSQSIKTSKADLAPTSKVLAKPTP 199
Db 181 TSLTVNKETTIVETKETTNTNKQTSDDGKEKTSKAKTSQSIKTSKADLAPTSKVLAKPTP 240
QY 200 KAETTTKGPALTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 259
Db 241 KAETTTKGPALTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 300
QY 260 KEPAPTTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 319
Db 301 KEPAPTTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 360
QY 320 TTPKEPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 379
Db 361 TTPKEPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 420
QY 380 PSPTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPKBPASTTTPK 439
```

Db 421 PSPTPKBPAPTTTSAPTTKEPAPTTTSAPTTKEPAPTTTKEPAPTTTKEPAPTTT 480  
Qy 440 KKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 499  
Db 481 KKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 540  
Qy 500 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 559  
Db 541 PTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 600  
Qy 560 KKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 619  
Db 601 KKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 660  
Qy 620 PKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 679  
Db 661 PKGTAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 720  
Qy 680 TSDKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 739  
Db 721 TSDKAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 780  
Qy 740 DESTPELSABPTPKALENSPKBPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPTTTA 799  
Db 781 DESTPELSABPTPKALENSPKBPGVPTTKTPAATKPEMTTAKDKTTERDLRTTPTTTA 840  
Qy 800 APRMTKETATTEKTTESKITATTQVSTTTQDTPPKITLTKTTTLPKVVTTTKKTIIT 859  
Db 841 APRMTKETATTEKTTESKITATTQVSTTTQDTPPKITLTKTTTLPKVVTTTKKTIIT 900  
Qy 860 TTEIMNKPBEAKPKDRATNSKATPKBOKPTKAPKPTSTKPKMVRVKPKTTPPTPR 919  
Db 901 TTEIMNKPBEAKPKDRATNSKATPKBOKPTKAPKPTSTKPKMVRVKPKTTPPTPR 960  
Qy 920 KMTSTWPELNPSTRIAEAMLQTTTRNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHV 979  
Db 961 KMTSTWPELNPSTRIAEAMLQTTTRNQTPNSKLVEVNPKSEDAAGAEGETPHMLLRPHV 1020  
Qy 980 FMEVTPDMVLPVNPQGIHNPMLS 1006  
Db 1021 FMEVTPDMVLPVNPQGIHNPMLS 1047

RESULT 4

US-07-757-022B-58  
; Sequence 58, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserik, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 58:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1049 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-58  
  
Query Match 99.7%; Score 5358.7; DB 4; Length 1049;  
Best Local Similarity 95.9%; Pred. No. 4.5e-167;  
Matches 1006; Conservative 0; Mismatches 0; Indels 43; Gaps 1;  
  
Qy 1 MAWKTLPIVLLLLSVFVIQVSSQELSCGRCFESFERGRCDCDAQCKYDKCCPDYE 60  
Db 1 MAWKTLPIVLLLLSVFVIQVSSQELSCGRCFESFERGRCDCDAQCKYDKCCPDYE 60  
Qy 61 SPCAE-----VKDKKNTKKK 77  
Db 61 SPCAEHVSSEHQESSSSSSSSSIWKIKSSKNSAANRELQKLKVKDKNKNKTKK 120  
Qy 78 PTPKPPVVDAGSLDNGDFKVTTPDTSTTQHNKSVSTSPKITTAKPINRPSLPNSDTS 137  
Db 121 PTPKPPVVDAGSLDNGDFKVTTPDTSTTQHNKSVSTSPKITTAKPINRPSLPNSDTS 180  
Qy 138 KETSLTVNKEVTIVETKETTNNKQSTGDKETTSKETSIEKTSKDLAPTSKVLAKP 197  
Db 181 KETSLTVNKEVTIVETKETTNNKQSTGDKETTSKETSIEKTSKDLAPTSKVLAKP 240  
Qy 198 TPKAETTTKGPALTTPKEPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPT 257  
Db 241 TPKAETTTKGPALTTPKEPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTTKSAPT 300  
Qy 258 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 317  
Db 301 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 360  
Qy 318 APTTKEPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 377  
Db 361 APTTKEPTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 420  
Qy 378 KEPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 437  
Db 421 KEPSPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 480  
Qy 438 TPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 497  
Db 481 TPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 540  
Qy 498 LAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 557  
Db 541 LAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 600  
Qy 558 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 617  
Db 601 TPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660  
Qy 618 TTPKGTAPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 677





QY 858 ITTTEIMNKPBETAKPKDRATNSKATTPKQKPTKABKKPTSTKPKTMRVRKPKTTPT 917  
DB 901 ITTTEIMNKPBETAKPKDRATNSKATTPKQKPTKABKKPTSTKPKTMRVRKPKTTPT 960  
QY 918 PRKMTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSEDAGGAEGETPHMLLRP 977  
DB 961 PRKMTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSEDAGGAEGETPHMLLRP 1020  
QY 978 HVFMEVPTDMDYLPRVNPQGIINPMLS 1006  
DB 1021 HVFMEVPTDMDYLPRVNPQGIINPMLS 1049

RESULT 6

US-07-757-022B-46  
; Sequence 46, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA: US 07/546,114  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA: US 07/390,901  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5951  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-46

Query Match 99.7%; Score 5358; DB 4; Length 1320;  
Best Local Similarity 95.3%; Pred. No. 6.3e-167;  
Matches 1006; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQELSKGRCFSEFERGECDCDAQCKKYDKCCPDYE 60  
|||||

DB 1 MAWKTLPIYLLLLSVFVIQQVSSQELSKGRCFSEFERGECDCDAQCKKYDKCCPDYE 60  
QY 61 SFCA-----EYKONK 70  
|||||  
DB 61 SFCAVHNPTSPSSKAPPAGSAGQTIKSTTKRSPKPNKKTKKVIESEITEVCONK 120  
QY 71 KNRTKKKTPKPPVVVDEAGSLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSL 130  
DB 121 KNRTKKKTPKPPVVVDEAGSLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSL 180  
QY 131 PPNSTSKETSLSLVNKKETTVETKTTTINKQTSIDGKEKTTSAKETOSLEKTSADLAPT 190  
DB 181 PPNSTSKETSLSLVNKKETTVETKTTTINKQTSIDGKEKTTSAKETOSLEKTSADLAPT 240  
QY 191 SKVLAKPTPKAETTTKGPALITPKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPT 250  
DB 241 SKVLAKPTPKAETTTKGPALITPKPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPT 300  
QY 251 TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 310  
DB 301 TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 360  
QY 311 PTTPEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 370  
DB 361 PTTPEPAPTTPKEPTPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 420  
QY 371 EPAPTTPKEPTTPKEPAPTTPKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 430  
DB 421 EPAPTTPKEPTTPKEPAPTTPKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 480  
QY 431 PKEPAPTTPKGPAPTTPKEPAPTTPKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 490  
DB 481 PKEPAPTTPKGPAPTTPKEPAPTTPKSAPTTTKSAPTTTKSAPTTTKSAPTTTKSAPTTTK 540  
QY 491 TTTPEKLAPTTPEKAPATTPPELAPTTPPEPTTPPEPAPTTPKAAAPNTPKEPAPT 550  
DB 541 TTTPEKLAPTTPEKAPATTPPELAPTTPPEPTTPPEPAPTTPKAAAPNTPKEPAPT 600  
QY 551 PKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTKBPST 610  
DB 601 PKEPAPTTPKEPAPTTPKETAPTTPKGTAPTTLKEPAPTTPKPKAPKELAPTTKBPST 660  
QY 611 TSDKAPATTPKGTAPTTPEPAPTTPKGPAPTTPKGTAPTTLKEPAPTTPKPKAPKELAP 670  
DB 661 TSDKAPATTPKGTAPTTPEPAPTTPKGPAPTTPKGTAPTTLKEPAPTTPKPKAPKELAP 720  
QY 671 TTTKGTSTTSKAPATTPKETAPTTPKGPAPTTPKGPAPTTPPEPTTPPEPTTPPEPTTP 730  
DB 721 TTTKGTSTTSKAPATTPKETAPTTPKGPAPTTPKGPAPTTPPEPTTPPEPTTPPEPTTP 780  
QY 731 EPTTIHKSPPDESTEPESAEPTPKALENSPKPEGVPTTKTPAAATKPEMTTTAKDKTTERDL 790  
DB 781 EPTTIHKSPPDESTEPESAEPTPKALENSPKPEGVPTTKTPAAATKPEMTTTAKDKTTERDL 840  
QY 791 RTTPETTTAAPKMTKETAITTEKTTESKITATTTQVSTTTQDTTTPPKITTLKTTILAPK 850  
DB 841 RTTPETTTAAPKMTKETAITTEKTTESKITATTTQVSTTTQDTTTPPKITTLKTTILAPK 900  
QY 851 VTTTKTITTTTEIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPKTSKKPKTMRVR 910  
DB 901 VTTTKTITTTTEIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPKTSKKPKTMRVR 960  
QY 911 KPKTTPTRKMTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSEDAGGAEGE 970  
DB 961 KPKTTPTRKMTSTMPNLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPXSEDAGGAEGE 1020  
QY 971 PHMLLRPHVEMPEVTPDMDYLPRVNPQGIINPMLS 1006  
DB 1021 PHMLLRPHVEMPEVTPDMDYLPRVNPQGIINPMLS 1056

; Sequence 60, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 60:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-60

Query Match 99.7%; Score 5358; DB 4; Length 1320;  
Best Local Similarity 95.3%; Pred. No. 6.3e-167;  
Matches 1006; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

QY 1 MAWKTLPIYLLILLISVFVIQVSSQELCKGRCFESFERGECDCDAOCKYDKCCPDYE 60  
DB 1 MAWKTLPIYLLILLISVFVIQVSSQELCKGRCFESFERGECDCDAOCKYDKCCPDYE 60  
QY 61 SFCA-----EVKDNK 70  
DB 61 SFCAEVHNFTSPSSKAPPSPGASQTIKSTTKRSPKPPNKKTKKVVIESEITEVKNK 120  
QY 71 KNRTKKKTPKPPVVDAGSLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSL 130  
DB 121 KNRTKKKTPKPPVVDAGSLDNGDFKVTTPDSTTQHNKVSTSPKITTAKPINRPSL 180  
QY 131 PPNDSKTSKTSLVNKEVTTETTKNTKOTSDTGKKTSAKETQSIKTSKADLAPT 190  
DB 181 PPNDSKTSKTSLVNKEVTTETTKNTKOTSDTGKKTSAKETQSIKTSKADLAPT 240  
QY 191 SKVLAKPTPKAETTTKGPALATTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPT 250

DB 241 SKVLAKPTPKAETTTKGPALATTPKEPTPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPT 300  
QY 251 TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 310  
DB 301 TTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 360  
QY 311 PTTPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 370  
DB 361 PTTPEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 420  
QY 371 EPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 430  
DB 421 EPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 480  
QY 431 PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 490  
DB 481 PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 540  
QY 491 TPTPEKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 550  
DB 541 TPTPEKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 600  
QY 551 PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 610  
DB 601 PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 660  
QY 611 TSDKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 670  
DB 661 TSDKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 720  
QY 671 TTTKGPTSTSDKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 730  
DB 721 TTTKGPTSTSDKAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPT 780  
QY 731 EPTTIHKSPDESTPELSAETPKALENSPKGCVPTTKTPAATKPEMTTTAKDKTTERDL 790  
DB 781 EPTTIHKSPDESTPELSAETPKALENSPKGCVPTTKTPAATKPEMTTTAKDKTTERDL 840  
QY 791 RTPTETTTAAPKMTKETATTTKTTESKITATTTQVTSITTTQDTTFFKITTLLKTTLLAPK 850  
DB 841 RTPTETTTAAPKMTKETATTTKTTESKITATTTQVTSITTTQDTTFFKITTLLKTTLLAPK 900  
QY 851 VTTTKKTTTITTEIMNKPEETAKPKDRATNSKATTPKPKOKPKAPKKPSTSKPKTMPRVR 910  
DB 901 VTTTKKTTTITTEIMNKPEETAKPKDRATNSKATTPKPKOKPKAPKKPSTSKPKTMPRVR 960  
QY 911 KPKTTPTPKMTSTMPELNPTSRIAEAMLOTTTRPNQTPNSKLVVNPKSESDAGGAGET 970  
DB 961 KPKTTPTPKMTSTMPELNPTSRIAEAMLOTTTRPNQTPNSKLVVNPKSESDAGGAGET 1020  
QY 971 PHMLLRPHVFMPEVTPDMDYLPVRPNQGIINPMLS 1006  
DB 1021 PHMLLRPHVFMPEVTPDMDYLPVRPNQGIINPMLS 1056

RESULT 8  
US-07-757-022B-52  
; Sequence 52, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.

ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1363 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-52

Query Match 99.6%; Score 5353.7; DB 4; Length 1363;  
Best Local Similarity 91.5%; Pred. No. 9.1e-167;  
Matches 100%; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSKGRCFESFERGECDCDAQCKKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLSVFVIQVSSQELSKGRCFESFERGECDCDAQCKKYDKCCPDYE 60  
QY 61 SFCAE----- 65  
Db 61 SFCAEVHNPTSPPSKKAPPSPGASQTIKSTTKRSPKPNKKTKKVIESEBITEBHSVS 120  
QY 66 -----VKDNKNRTKKKTPKPPVWDE 87  
Db 121 ENQESSSSSSSSSTIWKIKSSKNSAANRELQKKLVKDNKNRTKKKTPKPPVWDE 180  
QY 88 AGSGLDNGDPKVTTPDTSITTHQNKVSTSPKITTAKPINRPSLPNSDTSKETSLSVKNKE 147  
Db 181 AGSGLDNGDPKVTTPDTSITTHQNKVSTSPKITTAKPINRPSLPNSDTSKETSLSVKNKE 240  
QY 148 TTIVETKETTNNKQTSIDGKEKTTSAKETQSIBKTSKAKOLAPTSKVLAKTPKAEATTIKG 207  
Db 241 TTIVETKETTNNKQTSIDGKEKTTSAKETQSIBKTSKAKOLAPTSKVLAKTPKAEATTIKG 300  
QY 208 PALTTKPEPTTPKPEASTTKPEPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTT 267  
Db 301 PALTTKPEPTTPKPEASTTKPEPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTT 360  
QY 268 KEPAPTTKPEAPTTTKSAPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTTP 327  
Db 361 KEPAPTTKPEAPTTTKSAPTTPKPEAPTTTKKPAPTTPKPEAPTTTKPEPTTP 420  
QY 328 TTPKPEAPTTKPEAPTTKPGAPATPKKPAPTTPKPEAPTTTKPEAPTTTKESPTTPKE 387  
Db 421 TTPKPEAPTTKPEAPTTKPGAPATPKKPAPTTPKPEAPTTTKPEAPTTTKESPTTPKE 480

QY 388 PAPTTTKGAPTTTKPEAPTTTKSAPTTPKPEPTTPKPEAPTTTKPEAPTTTKPKKPAPTTP 447  
Db 481 PAPTTTKGAPTTTKPEAPTTTKSAPTTPKPEPTTPKPEAPTTTKPEAPTTTKPKKPAPTTP 540  
QY 448 KEPAPTTKPEAPTTTKKPAATAPKPEAPTTTKETAPTTPKLTPTTPEKLAPTTPEKPA 507  
Db 541 KEPAPTTKPEAPTTTKKPAATAPKPEAPTTTKETAPTTPKLTPTTPEKLAPTTPEKPA 600  
QY 508 PTTPEELAPTTPEEPTTPKPEAPTTTKAAANPTPKPEAPTTTKPEAPTTTKPEAPTTTP 567  
Db 601 PTTPEELAPTTPEEPTTPKPEAPTTTKAAANPTPKPEAPTTTKPEAPTTTKPEAPTTTP 660  
QY 568 KETAPTTKGTAPTTTLKPEAPTTPKKPAKELAPTTTKETPTSTTSKPKAPTTKGTAPTT 627  
Db 661 KETAPTTKGTAPTTTLKPEAPTTPKKPAKELAPTTTKETPTSTTSKPKAPTTKGTAPTT 720  
QY 628 PKPEAPTTKPEAPTTKGTAPTTTLKPEAPTTPKKPAKELAPTTTKGTSTTSKPKAPT 687  
Db 721 PKPEAPTTKPEAPTTKGTAPTTTLKPEAPTTPKKPAKELAPTTTKGTSTTSKPKAPT 780  
QY 688 TPKETAPTTKPEAPTTPKKPAPTTPETPPPTTSVSTPTTKEPTTIHKSPDESTEPELS 747  
Db 781 TPKETAPTTKPEAPTTPKKPAPTTPETPPPTTSVSTPTTKEPTTIHKSPDESTEPELS 840  
QY 748 ABPTPKALENSPKPGVPTTKTAAATKPEMTTTAKDKITERDLRTTPTTTAAPKMTKET 807  
Db 841 ABPTPKALENSPKPGVPTTKTAAATKPEMTTTAKDKITERDLRTTPTTTAAPKMTKET 900  
QY 808 ATTTEKTESKITATTQVTSSTTQDTPPKITTLKTTLAPKVTTTKTITTEIMNKP 867  
Db 901 ATTTEKTESKITATTQVTSSTTQDTPPKITTLKTTLAPKVTTTKTITTEIMNKP 960  
QY 868 BETAKPKDEATNSKATTPKPKQPKTAPKPKPTSTKPKETMPVRKPKTTPTPKMTSTMP 927  
Db 961 BETAKPKDEATNSKATTPKPKQPKTAPKPKPTSTKPKETMPVRKPKTTPTPKMTSTMP 1020  
QY 928 LNPTSRIBAMLQTTTTRPNQTPNSKLVEVNPKSEDAGGAEGTPEHMLLPHVFMPEVTPD 987  
Db 1021 LNPTSRIBAMLQTTTTRPNQTPNSKLVEVNPKSEDAGGAEGTPEHMLLPHVFMPEVTPD 1080  
QY 988 MDYLPRVFNQGIINPMLS 1006  
Db 1081 MDYLPRVFNQGIINPMLS 1099  
RESULT 9  
US-10-164-595-58  
; Sequence 58, Application US/10164595  
; Patent No. 6657054  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies, Inc  
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
; FILE REFERENCE: 1U 103 R1  
; CURRENT APPLICATION NUMBER: US/10/164,595  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 58  
; LENGTH: 1320  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-595-58  
Query Match 99.6%; Score 5349; DB 4; Length 1320;  
Best Local Similarity 95.1%; Pred. No. 1.2e-166;  
Matches 100%; Conservative 0; Mismatches 2; Indels 50; Gaps 1;  
QY 1 MAWKTLPIYLLLLSVFVIQVSSQELSKGRCFESFERGECDCDAQCKKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLLLSVFVIQVSSQELSKGRCFESFERGECDCDAQCKKYDKCCPDYE 60  
QY 61 SFCAE-----EYKDNK 70  
|||

Db 61 SFCAEHNPTSPSSKAPPPGASQTIKSTTKRSPKPPNNKKTKKVIESEEITEVDKNDK 120  
QY 71 KNRTKKKPTKPPVDEAGSLONGDFKVTTPDTSTTQHNVSTSPKLTAKINRPSL 130  
Db 121 KNRTKKKPTKPPVDEAGSLONGDFKVTTPDTSTTQHNVSTSPKLTAKINRPSL 180  
QY 131 PPNSTSKETSLVNVKETTVEKETTNNKQSTDGKEKTTSAKETQSIKTSKADLAPT 190  
Db 181 PPNSTSKETSLVNVKETTVEKETTNNKQSTDGKEKTTSAKETQSIKTSKADLAPT 240  
QY 191 SKVLAKPTKAEHTTKGALTTKAPTPTPKPASTTTPKEPTPTTKSAPTTPKSPAPT 250  
Db 241 SKVLAKPTKAEHTTKGALTTKAPTPTPKPASTTTPKEPTPTTKSAPTTPKSPAPT 300  
QY 251 TTKSAPTTPKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKP 310  
Db 301 TTKSAPTTPKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKP 360  
QY 311 PTPKSPAPTPTKEPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTK 370  
Db 361 PTPKSPAPTPTKEPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTK 420  
QY 371 EPAPTPTTKSPPTTKPAPTPTTKSAPTTTKPAPTPTTKSAPTTTKPAPTPTTKPAPTPT 430  
Db 421 EPAPTPTTKSPPTTKPAPTPTTKSAPTTTKPAPTPTTKSAPTTTKPAPTPTTKPAPTPT 480  
QY 431 PKPAPTPTTKKAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTK 490  
Db 481 PKPAPTPTTKKAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTKPAPTPTTK 540  
QY 491 TPTTPEKLAPTPEKAPTPTPELAPTPTPEPTPTTPEPAPTPTPKAAAPTTPKEPAPT 550  
Db 541 TPTTPEKLAPTPEKAPTPTPELAPTPTPEPTPTTPEPAPTPTPKAAAPTTPKEPAPT 600  
QY 551 PKPAPTPTTKPAPTPTTKAPTPTTKGAPTTLKEPAPTPTTKPAPTPTTKPAPTPTTK 610  
Db 601 PKPAPTPTTKPAPTPTTKAPTPTTKGAPTTLKEPAPTPTTKPAPTPTTKPAPTPTTK 660  
QY 611 TSDKAPTPTPKGTAPTPTKEPAPTPTKEPAPTPTKEPAPTPTKEPAPTPTKEPAPTPT 670  
Db 661 TCDKAPTPTPKGTAPTPTKEPAPTPTKEPAPTPTKEPAPTPTKEPAPTPTKEPAPTPT 720  
QY 671 TTTKGTSTSDKAPTPTKETAPTPTKEPAPTPTPKKAPTPTPEPTPTTSVSTPTTK 730  
Db 721 TTTKGTSTSDKAPTPTKETAPTPTKEPAPTPTPKKAPTPTPEPTPTTSVSTPTTK 780  
QY 731 EPTTIHKSDSTPBLASPTPKALENSPKGVPPTTKTPTAATKPEMTTAKDKTTERDL 790  
Db 781 EPTTIHKSDSTPBLASPTPKALENSPKGVPPTTKTPTAATKPEMTTAKDKTTERDL 840  
QY 791 RTTPTTTAAAPKWTETATTEKTTESKITATTQVSTTTQDTTPEKLTTLKTTTLAPK 850  
Db 841 RTTPTTTAAAPKWTETATTEKTTESKITATTQVSTTTQDTTPEKLTTLKTTTLAPK 900  
QY 851 VTTTKKTTITTTIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKTWPVR 910  
Db 901 VTTTKKTTITTTIMNKPEETAKPKDRATNSKATTPKQKPTKAPKPTSTKKPKTWPVR 960  
QY 911 KPKTPTPRKMTSTMPELNPTSRIAEMLQTTTRNQTPNSKLVNPKSEADGAGGET 970  
Db 961 KPKTPTPRKMTSTMPELNPTSRIAEMLQTTTRNQTPNSKLVNPKSEADGAGGET 1020  
QY 971 PHMLRPHVFMEVTPDMDYLRVFNQGIINPMLS 1006  
Db 1021 PHMLRPHVFMEVTPDMDYLRVFNQGIINPMLS 1056

## RESULT 10

US-07-757-022B-48

; Sequence 48, Application US/07757022B

; Patent No. 6433142

; GENERAL INFORMATION:

; APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseri, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1354 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-48

Query Match 99.5%; Score 5344.6; DB 4; Length 1354;

Best Local Similarity 92.3%; Pred. No. 1.8e-166;

Matches 1006; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQVSSQ-----25

Db 1 MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGRCGEGYSDATCNCNDYHMECCPDF 60

QY 26 -----ELSCGRCPESFERGECDCDAQCKYDKCCPDYDFSCAB-----65

Db 61 KRVCCTAELSCGRCPESFERGECDCDAQCKYDKCCPDYDFSCABEHSVSENQSSSSS 120

QY 66 -----VKDNKNRTKKKPTPKPPVVDGAGSLDNGD 96

Db 121 SSSSSSSTIWKIKSKNSAANRELQKKUKVNDKNRTKKKPTPKPPVVDGAGSLDNGD 180

QY 97 FKVTTPDTSTTQHNVSTSPKLTAKINRPSLPNNSDTSKETSLTVNKETTVEKETT 156

Db 181 FKVTTPDTSTTQHNVSTSPKLTAKINRPSLPNNSDTSKETSLTVNKETTVEKETT 240

QY 157 TTNKQSTSDGKEKTSKETSIAKDLAPTSKVLAKPTPKAETTTKGPALTTPKP 216

Db 241 TTNKQSTSDGKEKTSKETSIAKDLAPTSKVLAKPTPKAETTTKGPALTTPKP 300

QY 217 TPTPKPASTTTPKEPTPTTKSAPTTPKEPAPTTPKEPAPTTPKEPAPTTPK 276

301 TTTTTPKEASITPKXPTTIIKSAITPKKEPAPTTTKSAPTTPKKEPAPTTTKKEPAPTTTK 360  
277 EPAPTTTKEPAPTTTKSAPTTKEPAPTTTKKPAITPKKEPAPTTTKKEPAPTTTKKEPAPT 336  
361 EPAPTTTKEPAPTTTKSAPTTKEPAPTTTKKPAITPKKEPAPTTTKKEPAPTTTKKEPAPT 420  
337 TKEPAPTTKEPAPTAAPKPAITPKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKSA 396  
421 TKEPAPTTKEPAPTAAPKPAITPKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTKSA 480  
397 PTTTKEPAPTTTKSAPTTKEPAPTTTKKEPAPTTTKKEPAPTTTKKPAITPKKEPAPTTTK 456  
481 PTTTKEPAPTTTKSAPTTTKESAPTTTKKEPAPTTTKKEPAPTTTKKPAITPKKEPAPTTTK 540  
457 EPAPTTTKEPAPTAPEPAPTTTKETAPTTTKKLTPTTPEKLAITPKKEPAPTTTKPEELAP 516  
541 EPAPTTTKEPAPTAPEPAPTTTKETAPTTTKKLTPTTPEKLAITPKKEPAPTTTKPEELAP 600  
517 TTPPEPTPTTPEPAPTTTKAANPTPKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTK 576  
601 TTPPEPTPTTPEPAPTTTKAANPTPKKEPAPTTTKKEPAPTTTKKEPAPTTTKKEPAPTTTK 660  
577 GTAPTTTKEPAPTTKPAKPAITPKKEPAPTTTKETPTSTTSKDPAPTTTKGTAPTTTKKEPAPT 636  
661 GTAPTTTKEPAPTTKPAKPAITPKKEPAPTTTKETPTSTTSKDPAPTTTKGTAPTTTKKEPAPT 720  
637 KEAPPTTKGTAPTTKEPAPTTKPAKPAITPKKEPAPTTTKGTPTSTTSKDPAPTTTKGTAPT 696  
721 KEAPPTTKGTAPTTKEPAPTTKPAKPAITPKKEPAPTTTKGTPTSTTSKDPAPTTTKGTAPT 780  
697 PKEPAPTTKPAITPKKEPAPTTTSEVSTPTTKETPTTIHKSDESTPELSAPTTKPALE 756  
781 PKEPAPTTKPAITPKKEPAPTTTSEVSTPTTKETPTTIHKSDESTPELSAPTTKPALE 840  
757 NSPKPGVPTTKPAATKEMTTAKDKTERDLRTTPTTTAAPKMTKETATTTTEKTVE 816  
841 NSPKPGVPTTKPAATKEMTTAKDKTERDLRTTPTTTAAPKMTKETATTTTEKTVE 900  
817 SKIATTTQVSTTTQDTPFKITTLKTTTLAPKVTITTKITITTEIMNKPESTAKPKOR 876  
901 SKIATTTQVSTTTQDTPFKITTLKTTTLAPKVTITTKITITTEIMNKPESTAKPKOR 960  
877 AINSKATTKPKQPTKAPKPTSTKPKMTPRVRKPKTPTTKMTSTWPELNTSRIAE 936  
961 AINSKATTKPKQPTKAPKPTSTKPKMTPRVRKPKTPTTKMTSTWPELNTSRIAE 1020  
937 AMLQTTTRNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVFN 996  
1021 AMLQTTTRNQTPNSKLVEVNPKSEDAGGAEGETPHMLLRPHVFMPEVTPDMDYLPVFN 1080  
997 QGIIINPMLS 1006  
1081 QGIIINPMLS 1090

RESULT 11  
US-07-757-022B-40  
; Sequence 40, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Geener, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.

ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: G1 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 40:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1361 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-40  
Query Match 99.5%; Score 5343.9; DB 4; Length 1361;  
Best Local Similarity 91.7%; Pred. No. 1.9e-166;  
Matches 1006; Conservative 0; Mismatches 0; Indels 91; Gaps 2;  
QY 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25  
DB 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 60  
QY 26 -----ELCKGRCFESFERGECDCDAOCKYDKCCPDYESFCA----- 64  
DB 61 KRVCTAELSCGRCFESFERGECDCDAOCKYDKCCPDYESFCAEVHNDTSPSSKKAP 120  
QY 65 -----EVKDNKKNRTKKKTKPKFPVVDEAG 89  
DB 121 PPSGASQTIKSTTKRSPKPNKKTKVIESEBEITEVKDNKKNRTKKKTKPKFPVVDEAG 180  
QY 90 SGLDNGDFKVTPTDSTTHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETT 149  
DB 181 SGLDNGDFKVTPTDSTTHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETT 240  
QY 150 VETKETTINKQTSIDGKEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTKGPA 209  
DB 241 VETKETTINKQTSIDGKEKTTSAKETOSIEKTSAKDLAPTSKVLAKPTPKAETTKGPA 300  
QY 210 LTTTKEPTPTTKPEASTTTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 269  
DB 301 LTTTKEPTPTTKPEASTTTPKEPTPTTIKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKE 360  
QY 270 PAPTTPKEPAPTTTKPEAPTTTKSAPTTPKPEAPTTTKKPAITPKKEPAPTTTKKEPAPT 329  
DB 361 PAPTTPKEPAPTTTKPEAPTTTKSAPTTPKPEAPTTTKKPAITPKKEPAPTTTKKEPAPT 420  
QY 330 PKEPAPTTKEPAPTTTKPEAPTAAPKPAITPKPEAPTTTKPEAPTTTKKEPAPTTTKKEPA 389  
DB 421 PKEPAPTTKEPAPTTTKPEAPTAAPKPAITPKPEAPTTTKPEAPTTTKKEPAPTTTKKEPA 480







Db 781 TAPTLKBPAPPTPKKPAKELAPTTTKGPTSTTSKPAPTPKETAPTKBPAPTPK 840  
QY 707 KPAPTTPPTPTSTSVSTTTTKEPTTHKSPDESPELSABTPKALENSPKPGVPT 766  
Db 841 KPAPTTPPTPTSTSVSTTTTKEPTTHKSPDESPELSABTPKALENSPKPGVPT 900  
QY 767 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMTKETATTTKTTESKITATTQV 826  
Db 901 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMTKETATTTKTTESKITATTQV 960  
QY 827 TSTTTQDTTPFKITTLTKTTLAPKVTTKKTIITTEIMNKPEBTAKPKORATNSKATTPK 886  
Db 961 TSTTTQDTTPFKITTLTKTTLAPKVTTKKTIITTEIMNKPEBTAKPKORATNSKATTPK 1020  
QY 887 POKPTKAPKPTSTKPKMVRKPKTTPTRKMTSTWPELNPSTRIAEAMLOTTTREN 946  
Db 1021 POKPTKAPKPTSTKPKMVRKPKTTPTRKMTSTWPELNPSTRIAEAMLOTTTREN 1080  
QY 947 QTPNSKLVEVNPKSEDAGAGBETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1006  
Db 1081 QTPNSKLVEVNPKSEDAGAGBETPHMLLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1140

RESULT 14  
US-07-757-022B-62  
; Sequence 62, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Geener, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ceert, Luan  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1404 amino acids  
; TYPE: AMINO ACID

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-62  
Query Match 99.4%; Score 5339.6; DB 4; Length 1404;  
Best Local Similarity 88.2%; Pred. No. 2.7e-166;  
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;  
QY 1 MAWKTIPIYLLLLSVFVIQOVSSQ-----25  
Db 1 MAWKTLPIVLLLLSVFVIQOVSSQDLSSCAGRCGEGYSRDATCNCNDYNCQHMECCPDF 60  
QY 26 -----ELSCKGRCFESFERGECDCDAQCKKYDKCPDYESFCAE-----65  
Db 61 KRVTAEELCKGRCFESFERGECDCDAQCKKYDKCPDYESFCAE-----120  
QY 66 -----65  
Db 121 PPSGASQTIKSTTKRSPKPPNKKTKVIESEBI TEHSVSENQESSSSSSSSSSTIW 180  
QY 66 -----VKONKNRTKKKPTPKPPVVDEAGSLDNGDFKVTTPDTST 106  
Db 181 KIKSSKNSAANRELQKKLVKDKNKNRTKKKPTPKPPVVDEAGSLDNGDFKVTTPDTST 240  
QY 107 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTINKQSTDG 166  
Db 241 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTINKQSTDG 300  
QY 167 KEKTTISAKETQSIKTSKADLAPT SKVLAKPTPKAETTTKGPALTTPKPEPTTTPKEPAS 226  
Db 301 KEKTTISAKETQSIKTSKADLAPT SKVLAKPTPKAETTTKGPALTTPKPEPTTTPKEPAS 360  
QY 227 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTKSGAPTTTPKEPAPTTTPKEPAPTTTKPE 286  
Db 361 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTKSGAPTTTPKEPAPTTTPKEPAPTTTKPE 420  
QY 287 APTTTKSAPTTTPKEPAPTTPKPAPTTTPKEPAPTTPKPEPTTTPKEPAPTTTPKEPAPTTPK 346  
Db 421 APTTTKSAPTTTPKEPAPTTPKPAPTTTPKEPAPTTPKPEPTTTPKEPAPTTTPKEPAPTTPK 480  
QY 347 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 406  
Db 481 EPAPTAPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 540  
QY 407 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKKP 466  
Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKKP 600  
QY 467 APTAPKEPAPTTPKETAPTTPKKLTPTTPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT 526  
Db 601 APTAPKEPAPTTPKETAPTTPKKLTPTTPKELAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTT 660  
QY 527 PEEPAPTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 586  
Db 661 PEEPAPTTPKAAAPNTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 720  
QY 587 APTTPKAPKELAPTTTKETSTSTSDKAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 646  
Db 721 APTTPKAPKELAPTTTKETSTSTSDKAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 780  
QY 647 TAPTTLKEPAPTTPKKPAKELAPTTTKGPTSTTSKPAPTPKETAPTTTPKEPAPTTTPK 706  
Db 781 TAPTTLKEPAPTTPKKPAKELAPTTTKGPTSTTSKPAPTPKETAPTTTPKEPAPTTTPK 840  
QY 707 KPAPTTPPTPTSTSVSTPTTTTKEPTTHKSPDESPELSABTPKALENSPKPGVPT 766  
Db 841 KPAPTTPPTPTSTSVSTPTTTTKEPTTHKSPDESPELSABTPKALENSPKPGVPT 900  
QY 767 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMTKETATTTKTTESKITATTQV 826  
Db 901 TKTPAATKPEMTTAKDKTTERDLRTPTTAAAPKMTKETATTTKTTESKITATTQV 960  
QY 827 TSTTTQDTTPFKITTLTKTTLAPKVTTKKTIITTEIMNKPEBTAKPKORATNSKATTPK 886

```
Db 961 TSTTTQDTPPKITLTKLTLAPKVTITTKLITITTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 887 POKPTKAPKPTSTKKPKTMPVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRN 946
Db 1021 POKPTKAPKPTSTKKPKTMPVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRN 1080
QY 947 QTPNSKLVEVNPKSEDAGAGBETHMULLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1006
Db 1081 QTPNSKLVEVNPKSEDAGAGBETHMULLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1140

RESULT 15
US-09-298-970A-1
; Sequence 1, Application US/09298970A
; Patent No. 6743774
; GENERAL INFORMATION:
; APPLICANT: Jav, Gregory D.
; TITLE OF INVENTION: TRIBONECTINS
; FILE REFERENCE: 21486-026
; CURRENT APPLICATION NUMBER: US/09/298,970A
; CURRENT FILING DATE: 2001-06-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1404
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-298-970A-1

Query Match 99.4%; Score 5339.6; DB 4; Length 1404;
Best Local Similarity 88.2%; Pred. No. 2.7e-166;
Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60
QY 26 -----ELSCGRCFPSFGRGECDCDAQCKKYDKCCPDYESFCAE----- 65
Db 61 KRVTCAELSCGRCFPSFGRGECDCDAQCKKYDKCCPDYESFCAEVHNPTSPPSKKAP 120
QY 66 ----- 65
Db 121 PFGASQTIKSTTKRSKPNKKTKKVIESEBITBEHSVSNQESSSSSSSSSSSTIW 180
QY 66 -----VKDNKNRTKKKTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 106
Db 181 KIKSSKNSAANREBLQKLVKDNKNRTKKKTPKPPVVDVDEAGSGLDNGDFKVTTPDTST 240
QY 107 TQHNKUSTSPKITITAKPINRPSLPNSDTSKETSLSLVNKEITTVETKETITTNKQJSTDG 166
Db 241 TQHNKVSTSPKITITAKPINRPSLPNSDTSKETSLSLVNKEITTVETKETITTNKQJSTDG 300
QY 167 KEKTTSAKETQSTKTSKADLAPTSKVLAKPTPKAETTTKGPALITTPKEPTTTPKEPAS 226
Db 301 KEKTTSAKETQSTKTSKADLAPTSKVLAKPTPKAETTTKGPALITTPKEPTTTPKEPAS 360
QY 227 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTTPKEPAPTTTKPEPAPTTTKPEPAPTTTKPE 286
Db 361 TTPKEPTPTTIKSAPTTKPEPAPTTTKSAPTTTPKEPAPTTTKPEPAPTTTKPEPAPTTTKPE 420
QY 287 APPTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTTPKEPTTTPKEPAPTTKPEPAPTTTKPE 346
Db 421 APPTTKSAPTTKPEPAPTTTKKAPTTKPEPAPTTTPKEPTTTPKEPAPTTKPEPAPTTTKPE 480
QY 347 EPAPTKPEPAPTTKPEPAPTTTPKEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKPEPAPT 406
Db 481 EPAPTKPEPAPTTKPEPAPTTTPKEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTTKPEAPT 540
QY 407 TTKSAPTTKPEPSTTTTKPEPAPTTKGPAPTTTPKKAPAPTTPKGPAPTTTPKPEPAPTTTKKP 466
Db 541 TTKSAPTTKPEPSTTTTKPEPAPTTPKGPAPTTTPKKAPAPTTPKGPAPTTTPKPEPAPTTTKKP 600
```

```
QY 467 APTAPEKAPPTPKETAPTTPKKLPTTPEKLPATTPPEKAPPTPEELAPTTPEEBPTPT 526
Db 601 APTAPEKAPPTPKETAPTTPKKLPTTPEKLPATTPPEKAPPTPEELAPTTPEEBPTPT 660
QY 527 PPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTPKPAPTTTPKETAAPTTPKGTAPTTLKBP 586
Db 661 PPEPAPTTPKAAAPNTPKEPAPTTTPKEPAPTTPKPAPTTTPKETAAPTTPKGTAPTTLKBP 720
QY 587 APTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTTPKEPAPTTPKPAPTTPKG 646
Db 721 APTTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTTPKEPAPTTPKPAPTTPKG 780
QY 647 TAPTTTLKBPAPPTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTTPKEPAPTTPK 706
Db 781 TAPTTTLKBPAPPTPKKAPKELAPTTTKEPTSTTSKAPAPTTPKGTAPTTTPKEPAPTTPK 840
QY 707 KPAPTTPEPTPTTSEVSTPTTTKSPPTTIHKSPDSESTPELSAEPTPKALENSPKPESGVP 766
Db 841 KPAPTTPEPTPTTSEVSTPTTTKSPPTTIHKSPDSESTPELSAEPTPKALENSPKPESGVP 900
QY 767 TKTPAATKPEMTTAKDKTTERDLATTPETTTAAAPKMTKETATTTTETTESKITATTTOV 826
Db 901 TKTPAATKPEMTTAKDKTTERDLATTPETTTAAAPKMTKETATTTTETTESKITATTTOV 960
QY 827 TSTTTQDTPPKITLTKLTLAPKVTITTKLITITTEIMNKPEETAKPKDRATNSKATTPK 886
Db 961 TSTTTQDTPPKITLTKLTLAPKVTITTKLITITTEIMNKPEETAKPKDRATNSKATTPK 1020
QY 887 POKPTKAPKPTSTKKPKTMPVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRN 946
Db 1021 POKPTKAPKPTSTKKPKTMPVRKPKTTPTPRKMTSTMPELNPTSRIAEAMLQTTTRN 1080
QY 947 QTPNSKLVEVNPKSEDAGAGBETHMULLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1006
Db 1081 QTPNSKLVEVNPKSEDAGAGBETHMULLRPHVFMPEVTPDMDYLRVFNQGIINPMLS 1140
```

Search completed: October 13, 2004, 11:59:09  
Job time : 25.4873 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 102.349 Seconds  
(without alignments)  
3171.696 Million cell updates/sec

Title: SEQ1-E  
Perfect score: 5373  
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....DMDYLPRVNGIINPMLS 1006

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 135658 seqs, 322682953 residues

Total number of hits satisfying chosen parameters: 1356558

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	5373	100.0	1038	13	US-10-124-557-74	Sequence 74, Appl
2	5373	100.0	1270	13	US-10-124-557-44	Sequence 44, Appl
3	5358.9	99.7	1311	13	US-10-124-557-42	Sequence 42, Appl
4	5358.7	99.7	1049	13	US-10-124-557-58	Sequence 58, Appl
5	5358.7	99.7	1313	13	US-10-124-557-142	Sequence 142, Appl
6	5358	99.7	1320	13	US-10-124-557-46	Sequence 46, Appl
7	5358	99.7	1320	13	US-10-124-557-60	Sequence 60, Appl
8	5353.7	99.6	1363	13	US-10-124-557-52	Sequence 52, Appl
9	5344.6	99.5	1354	13	US-10-124-557-48	Sequence 48, Appl
10	5343.9	99.5	1361	13	US-10-124-557-40	Sequence 40, Appl
11	5339.6	99.4	1140	13	US-10-124-557-104	Sequence 104, Appl
12	5339.6	99.4	1404	9	US-09-802-207-30	Sequence 30, Appl
13	5339.6	99.4	1404	11	US-09-897-188-1	Sequence 1, Appli
14	5339.6	99.4	1404	13	US-10-124-557-2	Sequence 2, Appli

15 5339.6 99.4 1404 13 US-10-124-557-62 Sequence 62, Appl  
16 5241.8 97.6 1022 13 US-10-124-557-84 Sequence 84, Appl  
17 5232.6 97.4 1314 13 US-10-124-557-50 Sequence 50, Appl  
18 5011 93.3 941 13 US-10-124-557-14 Sequence 14, Appl  
19 1975.4 36.8 792 9 US-09-802-207-27 Sequence 27, Appl  
20 1409.9 26.2 538 14 US-10-038-694-3 Sequence 3, Appl  
21 1335.3 24.9 5179 9 US-09-922-217-1068 Sequence 1068, Ap  
22 1335.3 24.9 5179 9 US-09-833-263-1068 Sequence 1068, Ap  
23 1335.3 24.9 5179 13 US-10-025-380-1068 Sequence 121, App  
24 1335.3 24.9 5179 16 US-10-734-564-121 Sequence 4, Appli  
25 1146.1 21.3 292 16 US-10-468-910-4 Sequence 428, App  
26 932 17.3 1460 14 US-10-295-027-428 Sequence 35612, A  
27 920.2 17.1 1325 9 US-09-864-761-35612 Sequence 108, App  
28 892.3 16.6 1367 9 US-09-801-368-108 Sequence 5013, Ap  
29 845.4 15.7 6642 14 US-10-369-493-5013 Sequence 5784, Ap  
30 844.6 15.7 3507 14 US-10-369-493-5784 Sequence 2318, Ap  
31 834.6 15.5 2090 16 US-10-408-765A-2318 Sequence 41545, A  
32 832.7 15.3 697 15 US-10-425-114-41545 Sequence 79, Appl  
33 822.8 15.3 1151 10 US-09-825-751A-79 Sequence 5, Appli  
34 822 15.3 19723 15 US-10-084-846A-5 Sequence 71232, A  
35 811.2 15.1 3256 16 US-10-408-765A-174 Sequence 21, Appl  
36 811.2 15.1 3256 16 US-10-701-490-9 Sequence 10, Appl  
37 806.7 15.0 2397 15 US-10-282-122A-71232 Sequence 158, App  
38 803.4 15.0 3256 9 US-09-919-172-98 Sequence 311, App  
39 803.4 15.0 3256 10 US-09-919-039-21 Sequence 120, App  
40 798.4 14.9 1255 10 US-09-996-069-10 Sequence 5, Appli  
41 798.4 14.9 1255 14 US-10-171-311-158 Sequence 9, Appli  
42 798.4 14.9 1255 14 US-10-177-293-311 Sequence 98, Appl  
43 798.4 14.9 1255 16 US-10-734-564-120 Sequence 21, Appl  
44 793.6 14.8 22152 16 US-10-715-066-5 Sequence 10, Appl  
45 791.9 14.7 5935 14 US-10-243-243A-8 Sequence 5, Appli

## ALIGNMENTS

## RESULT 1

US-10-124-557-74  
; Sequence 74, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124.557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989



```

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQELSCCKGRCFESFERGECDCDAQCKKYDKCCPDYE 60
Db 1 MAWKTLPIYLLLLSVFVIQOVSSQELSCCKGRCFESFERGECDCDAQCKKYDKCCPDYE 60
QY 61 SFCAEVKDNKNNRKKKPTPKPPVVDAGSLDNGDFKVTTPDTSTTQHNKVSTSPKITT 120
Db 61 SFCAEVKDNKNNRKKKPTPKPPVVDAGSLDNGDFKVTTPDTSTTQHNKVSTSPKITT 120
QY 121 AKPINPRPSLPNSDTSKETSILVNKETTIVETKETTNNKQSTSDGKEKTTSAKETOSIE 180
Db 121 AKPINPRPSLPNSDTSKETSILVNKETTIVETKETTNNKQSTSDGKEKTTSAKETOSIE 180
QY 181 KTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPASTTPKEPTTIKSA 240
Db 181 KTSAKDLAPTSKVLAKPTPKAETTTKGPALTTPKEPTTTPKEPASTTPKEPTTIKSA 240
QY 241 PTPKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTPKE 300
Db 241 PTPKPEPAPTTTKSAPTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKPEPAPTTTKSAPTTPKE 300
QY 301 PAPTTPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTTTKSAPTTTKPEPSP 420
Db 301 PAPTTPKKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTTTKSAPTTTKPEPSP 420
QY 421 TTTKPEPAPTTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTTTKKAPAPKAPPTPK 480
Db 421 TTTKPEPAPTTTPKEPAPTTTPKKPAPTTTPKEPAPTTTPKEPAPTTTKKAPAPKAPPTPK 480
QY 481 ETAPPTPKLTPPTPEKLAPTTPKEPAPTTPEELAPTTPEEPPTTPPEPAPTTTPKAAAP 540
Db 481 ETAPPTPKLTPPTPEKLAPTTPKEPAPTTPEELAPTTPEEPPTTPPEPAPTTTPKAAAP 540
QY 541 NTPKPEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTTPKGTAPTTTPKAPKAPKELA 600
Db 541 NTPKPEPAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTTPKGTAPTTTPKAPKAPKELA 600
QY 601 PTTTKEPTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTATTTTPKEPAPTT 660
Db 601 PTTTKEPTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTATTTTPKEPAPTT 660
QY 661 KKPAPKELAPTTTKGFTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKAPKAPPTST 720
Db 661 KKPAPKELAPTTTKGFTSTSDKAPTTTPKGTAPTTTPKEPAPTTTPKAPKAPPTST 720
QY 721 SEVSTPTTKBPTTHKSPDESTPELSAPPTPKALENSKPEGVPTTKTPAATKPEMITT 780
Db 721 SEVSTPTTKBPTTHKSPDESTPELSAPPTPKALENSKPEGVPTTKTPAATKPEMITT 780
QY 781 AKDKTTERDLRTPTTTAAAPKWTKETATTTTEKTTESKITATTTQVSTTTQDTTPPKIT 840
Db 781 AKDKTTERDLRTPTTTAAAPKWTKETATTTTEKTTESKITATTTQVSTTTQDTTPPKIT 840
QY 841 TLKTTTLAPKVTTKKTTITTTIMNKPBTAKPKDRATNSKATTPKPKPTKAPKPTST 900
Db 841 TLKTTTLAPKVTTKKTTITTTIMNKPBTAKPKDRATNSKATTPKPKPTKAPKPTST 900
QY 901 KKPKTMPVRKPKTTPTRKMTSNMPELNPTSRIAEAMLQTTTRNQTPNSKLVENPKS 960
Db 901 KKPKTMPVRKPKTTPTRKMTSNMPELNPTSRIAEAMLQTTTRNQTPNSKLVENPKS 960
QY 961 EDAGGAEGETHMLLRPHVPEVTPDMDYLPVNPQGIINPMLS 1006
Db 961 EDAGGAEGETHMLLRPHVPEVTPDMDYLPVNPQGIINPMLS 1006

```

RESULT 3

US-10-124-557-42

; Sequence 42, Application US/10124557

; Publication No. US20020137894A1

```

;
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
;
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
;
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-10-124-557-42

```

```

Query Match          99.7%; Score 5358.9; DB 13; Length 1311;
Best Local Similarity 96.1%; Pred. No. 5.3e-134;
Matches 1006; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

```

```

QY 1 MAWKTLPIYLLLLSVFVIQOVSSQ----- 25
Db 1 MAWKTLPIYLLLLSVFVIQOVSSQDLSSCAGRCGYSRDATCNCDVNOHYMECCPDF 60
QY 26 -----ELSCCKGRCFESFERGECDCDAQCKKYDKCCPDYEFCAEVKDNKNNRKKKPT 79
Db 61 KRVCCTAELSCCKGRCFESFERGECDCDAQCKKYDKCCPDYEFCAEVKDNKNNRKKKPT 120
QY 80 KPFPVDEAGSLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPNSDTSKE 139
Db 121 KPFPVDEAGSLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRPSLPNSDTSKE 180
QY 140 TSLVNNKETTIVETKETTNNKQSTSDGKEKTTSAKETOSIEKTSKVLAKPTTP 199
Db 181 TSLVNNKETTIVETKETTNNKQSTSDGKEKTTSAKETOSIEKTSKVLAKPTTP 240
QY 200 KAETTTKGPALTTPKEPTTTPKEPASTTPKEPTTIKSAPTTPKAPPTTKSAPTTP 259
Db 241 KAETTTKGPALTTPKEPTTTPKEPASTTPKEPTTIKSAPTTPKAPPTTKSAPTTP 300

```

RESULT 4  
US-10-124-557-58  
; Sequence 58, Application US/10124557  
; Publication NO. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gsener, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:

Query Match	99.7%;	Score 5358.7;	DB 13;	Length 1049;
Best Local Similarity	95.9%;	Pred. No. 4.2e-134;		
Matches 1006;	Conservative 0;	Mismatches 0;	Indels 43;	Gaps 1;
QY	1	MAWKTLPIYLLLLSVFVIQYVSSQELSCGKRCFSEFERGECDCDAQCKYDKCCPDYE	60	
Db	1	MAWKTLPIYLLLLSVFVIQYVSSQELSCGKRCFSEFERGECDCDAQCKYDKCCPDYE	60	
QY	61	SFCAE-----VKNKKNRITKKK	77	
Db	61	SFCAEHSYSENGESSSSSSSSSSSSSIWKIKSSKNSAANRELQKKVKXDNKKNRITKKK	120	
QY	78	PTPKPPVVDAGSLDNGFKVTTPTSTTOHNKVSTSPKITTAKPINRPSLPNSDTS	137	
Db	121	PTPKPPVVDAGSLDNGFKVTTPTSTTOHNKVSTSPKITTAKPINRPSLPNSDTS	180	
QY	138	KETSLTVNKKETTVEKTTTNNKQTSDDGKEKTTSAKETQSIKTSAKOLAPTSKVLAKP	197	
Db	181	KETSLTVNKKETTVEKTTTNNKQTSDDGKEKTTSAKETQSIKTSAKDLAPTSKVLAKP	240	
QY	198	TPKAETTTKGPALTTTPKEBPTTPPKBPASTTPKEBPTTTIKSAPTTPKBPATTTKSAPT	257	
Db	241	TPKAETTTKGPALTTTPKEBPTTPPKBPASTTPKEBPTTTIKSAPTTPKBPATTTKSAPT	300	
QY	258	TPKEPAPTTTKBPAPTTPKBPATTTKBPATTTKGAAPTTPKBPAPTTPKKPAPTTPKP	317	
Db	301	TPKEPAPTTTKBPAPTTPKBPATTTKBPATTTKGAAPTTPKBPAPTTPKKPAPTTPKP	360	
QY	318	APTTTPKEBPTTPPKBPATTTKBPAPTTPKBPAPTTAPKKAPATTPPKBPAPTTTPKEBAPT	377	
Db	361	APTTTPKEBPTTPPKBPATTTKBPAPTTPKBPAPTTAPKKAPATTPPKBPAPTTTPKEBAPT	420	
QY	378	KEBSPPTPKBPATTTKSAPTTTKBPATTTKSAPTTPPKBPSPTTTKEBAPTTTPKEBAPT	437	
Db	421	KEBSPPTPKBPATTTKSAPTTTKBPATTTKSAPTTPPKBPSPTTTKEBAPTTTPKEBAPT	480	
QY	438	TPKKPAPTTPKBPATTPKBPATTTTKKPAPTAPEKBPATTPKETATPTPKLPTTPEK	497	
Db	491	TPKKPAPTTPKBPATTPKBPATTTTKKPAPTAPEKBPATTPKETATPTPKLPTTPEK	540	

198	Qy	TPKAEATTTKGPALTTTPKEPTPTTPKEBPASTTPKEPTPTTTKSAPTPTPKBPATTTPKSAPT	251
241	Db	TPKAEATTTKGPALTTTPKEPTPTTPKEBPASTTPKEPTPTTTKSAPTPTPKBPATTTPKSAPT	300
258	Qy	TPKEPATTTTKBPATTTTPKEPATTTTKBPATTTTKSAPTPTPKBPATTTPKKPAPATTTPEK	317
301	Db	TPKEPATTTTKBPATTTTPKEPATTTTKBPATTTTKSAPTPTPKBPATTTPKKPAPATTTPEK	360
318	Qy	APTTPKEPTPTTPKBPATTTTPKEPATTTTPKEBPATAPKKPATPTTPKEBPATTTTPKEPATTTT	377
361	Db	APTTPKEPTPTTPKBPATTTTPKEPATTTTPKEBPATAPKKPATPTTPKEBPATTTTPKEPATTTT	420
378	Qy	KEPSPPTTPKEPATTTTKSAPTTTTKBPATTTTKSAPTTPKEBPSTTTTKBPATTTTPKEBPAT	437
421	Db	KEPSPPTTPKEPATTTTKSAPTTTTKBPATTTTKSAPTTPKEBPSTTTTKBPATTTTPKEBPAT	480
438	Qy	TPKKPATPTTPKBPATTTTPKEPATTTTKKPAPATAPKBPATTTTPKETAPTTPKKLTPTTTPEK	497
481	Db	TPKKPATPTTPKBPATTTTPKEPATTTTKKPAPATAPKBPATTTTPKETAPTTPKKLTPTTTPEK	540

```
QY 498 LAPPTPEKPAATPEELAPPTPEBPTTTPKEPAATTPKAAAPNTPEPAATTPKEPAAT 557
Db 541 LAPPTPEKPAATPEELAPPTPEBPTTTPKEPAATTPKAAAPNTPEPAATTPKEPAAT 600
QY 558 TPKEPAATTPKEPAATTPKGTAPTTLKEPAATTPKKAPKELAPTTTKEPTSTTSKDPAP 617
Db 601 TPKEPAATTPKEPAATTPKGTAPTTLKEPAATTPKKAPKELAPTTTKEPTSTTSKDPAP 660
QY 618 TTPKGTAPTTPEKPAATTPKEPAATTPKGTAPTTLKEPAATTPKKAPKELAPTTTKEPT 677
Db 661 TTPKGTAPTTPEKPAATTPKEPAATTPKGTAPTTLKEPAATTPKKAPKELAPTTTKEPT 720
QY 678 STTSKDPAPTPKETAATTPKEPAATTPKGTAPTTLKEPAATTPPETPPPTTSEVSTPTTKEPTIHK 737
Db 721 STTSKDPAPTPKETAATTPKEPAATTPKGTAPTTLKEPAATTPPETPPPTTSEVSTPTTKEPTIHK 780
QY 738 SPDESPELSAETPKALENSPKRGVPTTKTAAATKPEMTTAKOKTTERDLRTTPETT 797
Db 781 SPDESPELSAETPKALENSPKRGVPTTKTAAATKPEMTTAKOKTTERDLRTTPETT 840
QY 798 TAAPKMTKETATTTETKTESKITATTQVTSSTTTQDTPPKITLTKTTLLAPKVTTTKT 857
Db 841 TAAPKMTKETATTTETKTESKITATTQVTSSTTTQDTPPKITLTKTTLLAPKVTTTKT 900
QY 858 ITTETIMNKPEETAKPKDRATNSKATTPKPKQKPTKAPKPKPTSTKPKTMRVRKPKTTPT 917
Db 901 ITTETIMNKPEETAKPKDRATNSKATTPKPKQKPTKAPKPKPTSTKPKTMRVRKPKTTPT 960
QY 918 PKMTSTMPELAPTSIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRP 977
Db 961 PKMTSTMPELAPTSIAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRP 1020
QY 978 HVFMPVETDMDYLPRVNPQGIINPMLS 1006
Db 1021 HVFMPVETDMDYLPRVNPQGIINPMLS 1049

RESULT 5
US-10-124-557-142.
; Sequence 142, Application US/10124557
; Publication No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
```

```
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 142:
US-10-124-557-142
```

```
Query Match 99.7%; Score 5358.7; DB 13; Length 1313;
Best Local Similarity 95.9%; Pred. No. 5,4e-134;
Matches 1006; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

QY 1 MAWKTLPIVILLLLSVFVIQQVSSQELSCCKGRCFESFERGREGCCDAQCKKYDKCCPDYE 60
Db 1 MAWKTLPIVILLLLSVFVIQQVSSQELSCCKGRCFESFERGREGCCDAQCKKYDKCCPDYE 60
QY 61 SFCAE-----VKDNKKNRKKK 77
Db 61 SFCAEHSHSVSEQESSSSSSSSSSSIWKIKSSKNSAANRELQKKLVKDNKKNRKKK 120
```

```
QY 78 PTPKPPVDEAGSGLDNGDFKVTTPDTSSTOHNKVSTSPKITTAKINPRPSLPNSDTS 137
Db 121 PTPKPPVDEAGSGLDNGDFKVTTPDTSSTOHNKVSTSPKITTAKINPRPSLPNSDTS 180
QY 138 KETSLTVNKETTVEKETTINKQTSQIEKTSKAKETOSIEKTSKAKOLAPTSKVLAKP 197
Db 181 KETSLTVNKETTVEKETTINKQTSQIEKTSKAKETOSIEKTSKAKOLAPTSKVLAKP 240
QY 198 TPKAETTTKGPALTTPEKPTTTPKEPAATTPKPEASTTTPKEPTTTIKSAPTTPEPAATTPKSAPT 257
Db 241 TPKAETTTKGPALTTPEKPTTTPKEPAATTPKPEASTTTPKEPTTTIKSAPTTPEPAATTPKSAPT 300
QY 258 TPKEPAATTTKEPAATTPKEPAATTTKEPAATTTKSAPTTTPKEPAATTPPKKAPATTPKEP 317
Db 301 TPKEPAATTTKEPAATTPKEPAATTTKEPAATTTKSAPTTTPKEPAATTPPKKAPATTPKEP 360
QY 318 APPTPEKPTTTPKEPAATTPKEPAATTPKAPATTPKAPATTPKPEPAATTPKEPAATTT 377
Db 361 APPTPEKPTTTPKEPAATTPKEPAATTPKAPATTPKAPATTPKPEPAATTPKEPAATTT 420
QY 378 KEPSPTTPEKPAATTTKSAPTTTPKEPAATTTKSAPTTTPKEPSPTTTPKEPAATTPKEPAAT 437
Db 421 KEPSPTTPEKPAATTTKSAPTTTPKEPAATTTKSAPTTTPKEPSPTTTPKEPAATTPKEPAAT 480
QY 438 TPKEPAATTPKEPAATTPKEPAATTTKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPKE 497
Db 481 TPKEPAATTPKEPAATTPKEPAATTTKAPATTPKAPATTPKAPATTPKAPATTPKAPATTPKE 540
QY 498 LAPTTPEKPAATTPPELAPTTPEBPTTTPKEPAATTPKAAAPNTPEPAATTPKEPAAT 557
Db 541 LAPTTPEKPAATTPPELAPTTPEBPTTTPKEPAATTPKAAAPNTPEPAATTPKEPAAT 600
QY 558 TPKEPAATTPKETAATTPKGTAPTTLKEPAATTPKKAPKELAPTTTKEPTSTTSKDPAP 617
Db 601 TPKEPAATTPKETAATTPKGTAPTTLKEPAATTPKKAPKELAPTTTKEPTSTTSKDPAP 660
QY 618 TTPKGTAPTTPEKPAATTPKEPAATTPKGTAPTTLKEPAATTPPKAPKELAPTTTKEPT 677
Db 661 TTPKGTAPTTPEKPAATTPKEPAATTPKGTAPTTLKEPAATTPPKAPKELAPTTTKEPT 720
QY 678 STTSKDPAPTPKETAATTPKEPAATTPKGTAPTTTPEPTTTPPTTSEVSTPTTKEPTIHK 737
Db 721 STTSKDPAPTPKETAATTPKEPAATTPKGTAPTTTPEPTTTPPTTSEVSTPTTKEPTIHK 780
```

QY 738 SPDESTPBLSEPTKALENSKPEGVPTTKTPAATKPEMTTAKDKTERDLRTTPTT 797  
Db 781 SPDESTPBLSEPTKALENSKPEGVPTTKTPAATKPEMTTAKDKTERDLRTTPTT 840  
QY 798 TAAPKMTETATTTKTESKITATTTOVSTTTQDTPFKITTLTKTTLAPKVTITTKT 857  
Db 841 TAAPKMTETATTTKTESKITATTTOVSTTTQDTPFKITTLTKTTLAPKVTITTKT 900  
QY 858 ITTTEIMNKPEETAKPKORATNSKATTPKPKQPTKAPKKPTSTKKPKMPVRKPKTTPT 917  
Db 901 ITTTEIMNKPEETAKPKORATNSKATTPKPKQPTKAPKKPTSTKKPKMPVRKPKTTPT 960  
QY 918 PKMTSTWPELNPTRIAEAMLOTTTRPNQTPNSKLVEVNPKSEADAGAGETPHMLLRP 977  
Db 961 PKMTSTWPELNPTRIAEAMLOTTTRPNQTPNSKLVEVNPKSEADAGAGETPHMLLRP 1020  
QY 978 HVFMEVTPDMDYLPVNPQGIINPMLS 1006  
Db 1021 HVFMEVTPDMDYLPVNPQGIINPMLS 1049

RESULT 6

US-10-124-557-46  
; Sequence 46, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/124,557  
; FILING DATE: 16-Apr-2002  
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 876-1170  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 46:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1320 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 46:  
US-10-124-557-46

Query Match 99.7%; Score 5358; DB 13; Length 1320;  
Best Local Similarity 95.3%; Pred. No. 5, 6e-134;  
Matches 1006; Conservative 0; Mismatches 0; Indels 50; Gaps 1;  
QY 1 MAWKTLPIYLLILLLSVFIQVSSQELSCGRCFESFERGRCDCAQCKKYDKCCPDYE 60  
Db 1 MAWKTLPIYLLILLLSVFIQVSSQELSCGRCFESFERGRCDCAQCKKYDKCCPDYE 60  
QY 61 SPCA-----EVDKDK 70  
Db 61 SPCAETHNPTSPSSKAPPPSGASQTIKSTTKRSPKPNKKTKKXVIESEITEVDKDK 120  
QY 71 KNRTKKPTPKPPVDEAGSLDNGDFKVTTPDTSTTOHKNKYSTSPKITTAKPINEPRL 130  
Db 121 KNRTKKPTPKPPVDEAGSLDNGDFKVTTPDTSTTOHKNKYSTSPKITTAKPINEPRL 180  
QY 131 PPNSTSKETSITVNKETTIVETKETTNNKQSTDGKETTSAKETQSEKTSADKLAPT 190  
Db 181 PPNSTSKETSITVNKETTIVETKETTNNKQSTDGKETTSAKETQSEKTSADKLAPT 240  
QY 191 SKVLAKPTPKAETTTKGPALITTPKBPPTTPKPEASTTPEKPTPTTIKSAPTPKPEAPT 250  
Db 241 SKVLAKPTPKAETTTKGPALITTPKBPPTTPKPEASTTPEKPTPTTIKSAPTPKPEAPT 300  
QY 251 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTTPKKPA 310  
Db 301 TTKSAPTTKPEAPTTTKPEAPTTTKPEAPTTTKPEAPTTTKSAPTTTKPEAPTTPKKPA 360  
QY 311 PTTPKPEAPTTKPEPTTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKK 370  
Db 361 PTTPKPEAPTTKPEPTTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTTKK 420  
QY 371 EPAPTTKPEPTTTKPEAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTT 430  
Db 421 EPAPTTKPEPTTTKPEAPTTTKSAPTTTKPEAPTTTKSAPTTTKPEAPTTTKPEAPTT 480  
QY 431 KPEAPTTPKKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPKETAPT 490  
Db 481 KPEAPTTPKKAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTPKETAPT 540  
QY 491 TPTTPEKLAPTTPEKAPTTPELAPTTPEEPTTPEEPTTPEEPTTPEEPTTPEEPTT 550  
Db 541 TPTTPEKLAPTTPEKAPTTPELAPTTPEEPTTPEEPTTPEEPTTPEEPTTPEEPTT 600  
QY 551 KPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 610  
Db 601 KPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 660  
QY 611 TSDKAPTTPKGTAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 670  
Db 661 TSDKAPTTPKGTAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTTKPEAPTT 720  
QY 671 TTTKGPTSTTSDKAPTTPKETAPTTPKPEAPTTPKKAPTTPKKAPTTPEPTPTTSEVST 730  
Db 721 TTTKGPTSTTSDKAPTTPKETAPTTPKPEAPTTPKKAPTTPKKAPTTPEPTPTTSEVST 780  
QY 731 EPTTIHKSDESTPELSAETPKALENSKPEGVPTTKTPAATKPEMTTAKDKTERDL 790  
Db 781 EPTTIHKSDESTPELSAETPKALENSKPEGVPTTKTPAATKPEMTTAKDKTERDL 840  
QY 791 RTTPTTTAAAPKMTKETATTTKTESKITATTTOVSTTTQDTPFKITTLTKTTLAPK 850  
Db 841 RTTPTTTAAAPKMTKETATTTKTESKITATTTOVSTTTQDTPFKITTLTKTTLAPK 900  
QY 851 VTTTKKTIITTEIMNKPEETAKPKORATNSKATTPKPKQPTKAPKKPTSTKKPKMPVR 910  
Db 901 VTTTKKTIITTEIMNKPEETAKPKORATNSKATTPKPKQPTKAPKKPTSTKKPKMPVR 960  
QY 911 KPKTTPTRKMTSTMPBLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEADAGAGET 970  
Db 961 KPKTTPTRKMTSTMPBLNPTSRIAEAMLOTTTRPNQTPNSKLVEVNPKSEADAGAGET 1020



Qy 971 PHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1006  
 Db 1021 PHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1056

RESULT 7  
 US-10-124-557-60  
 ; Sequence 60, Application US/10124557  
 ; Publication No. US20020137894A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Turner, Katherine  
 ; Clark, Stephen C.  
 ; Jacobs, Kenneth  
 ; Hewick, Rodney M.  
 ; Gesner, Thomas G.  
 ;  
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 ; NUMBER OF SEQUENCES: 143  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genetics Institute, Inc.  
 ; STREET: 87 CambridgePark Drive  
 ; CITY: Cambridge  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02140  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ;  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/124,557  
 ; FILING DATE: 16-Apr-2002  
 ; CLASSIFICATION: <Unknown>  
 ;  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/643,502  
 ; FILING DATE: 18-JAN-1991  
 ; APPLICATION NUMBER: US 07/546,114  
 ; FILING DATE: 29-JUN-1990  
 ; APPLICATION NUMBER: US 07/457,196  
 ; FILING DATE: 29-DEC-1989  
 ; APPLICATION NUMBER: US 07/390,901  
 ; FILING DATE: 08-AUG-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Cseri, Luann  
 ; REGISTRATION NUMBER: 31,822  
 ; REFERENCE/DOCKET NUMBER: GI 5190  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617)876-1170  
 ; TELEFAX: (617)876-5851  
 ;  
 ; INFORMATION FOR SEQ ID NO: 60:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1320 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ;  
 ; MOLECULE TYPE: protein  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 60:  
 US-10-124-557-60

Query Match 99.7%; Score 5358; DB 13; Length 1320;  
 Best Local Similarity 95.3%; Pred. No. 5.6e-134;  
 Matches 1006; Conservative 0; Mismatches 0; Indels 50; Gaps 1;

Qy 1 MAWKTLPIYLLLLLSVFVIQQVSSQELSCGRCFESFERGREGCDCAQCKYDKCCPYE 60  
 Db 1 MAWKTLPIYLLLLLSVFVIQQVSSQELSCGRCFESFERGREGCDCAQCKYDKCCPYE 60  
 Qy 61 SPCA-----EVKDK 70  
 Db 61 SPCA-----EVKDK 70  
 Qy 71 KNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDSTTQHNVKYSTSPKITTAKPINRPSL 130  
 Db 121 KNRTKKKTPKPPVVDGAGSLDNGDFKVTTPDSTTQHNVKYSTSPKITTAKPINRPSL 180

Qy 131 PPNSDTSKETSLSLVNKEETTIVETKTTTTNNKQTSIDGKEKTSIAKETOSIEKTSAKDLAPT 190  
 Db 181 PPNSDTSKETSLSLVNKEETTIVETKTTTTNNKQTSIDGKEKTSIAKETOSIEKTSAKDLAPT 240  
 Qy 191 SKVLAKPTPKAETTTKGPALTTTKEPTPTTKEPASTTTPKEPTTTTISAPTTTKEPAPT 250  
 Db 241 SKVLAKPTPKAETTTKGPALTTTKEPTPTTKEPASTTTPKEPTTTTISAPTTTKEPAPT 300  
 Qy 251 TTKSAPTTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 310  
 Db 301 TTKSAPTTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 360  
 Qy 311 PTTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 370  
 Db 361 PTTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 420  
 Qy 371 EPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 430  
 Db 421 EPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 480  
 Qy 431 PKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 490  
 Db 481 PKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 540  
 Qy 491 TPTTPEKLAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 550  
 Db 541 TPTTPEKLAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 600  
 Qy 551 PKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 610  
 Db 601 PKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 660  
 Qy 611 TSDKPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 670  
 Db 661 TSDKPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 720  
 Qy 671 TTTKGPTSTSDKPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 730  
 Db 721 TTTKGPTSTSDKPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPTPTTKEPAPT 780  
 Qy 731 EPTTIHKSPDESPELSAEPDPALENSPKEGVPTTKTAAATKPEMTTAKDKTTTERDL 790  
 Db 781 EPTTIHKSPDESPELSAEPDPALENSPKEGVPTTKTAAATKPEMTTAKDKTTTERDL 840  
 Qy 791 RTTPETTTAAAPKMTKETATTTTEKTTESKITATTTQVTSITTTQDTTTPFKITTLKTTTLAPK 850  
 Db 841 RTTPETTTAAAPKMTKETATTTTEKTTESKITATTTQVTSITTTQDTTTPFKITTLKTTTLAPK 900  
 Qy 851 VTTTKTITTTTBMNKPEETAKPKORATNSKATTPKQKPTKAPKPTSTKPKTMPVRV 910  
 Db 901 VTTTKTITTTTBMNKPEETAKPKORATNSKATTPKQKPTKAPKPTSTKPKTMPVRV 960  
 Qy 911 KPKTTPTPKMTSTMPELNPTSRIAEAMLOTTTRNQTNSKLVVNPKSEBAGAGET 970  
 Db 961 KPKTTPTPKMTSTMPELNPTSRIAEAMLOTTTRNQTNSKLVVNPKSEBAGAGET 1020  
 Qy 971 PHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1006  
 Db 1021 PHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1056

RESULT 8  
 US-10-124-557-52  
 ; Sequence 52, Application US/10124557  
 ; Publication No. US20020137894A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Turner, Katherine  
 ; Clark, Stephen C.  
 ; Jacobs, Kenneth  
 ; Hewick, Rodney M.  
 ; Gesner, Thomas G.  
 ;  
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

```

;
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA: US/10/124,557
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
;
; CLASSIFICATION DATA: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
;
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-10-124-557-52

Query Match 99.6%; Score 5353.7; DB 13; Length 1363;
Best Local Similarity 91.5%; Pred. No. 7.6e-134;
Matches 1006; Conservative 0; Mismatches 0; Indels 93; Gaps 1;

QY 1 MAWKTLPIYLLLSVFVIQVSSQELSKGRCFESPERGECDCDAQCKKYDKCCPDYE 60
DB 1 MAWKTLPIYLLLSVFVIQVSSQELSKGRCFESPERGECDCDAQCKKYDKCCPDYE 60
QY 61 SFCAE----- 65
DB 61 SFCAEHNPTSPSSKAPPSPGASQTIKSTTKRSPKPNKKTKKVIESEITEHSVS 120
QY 66 -----VKDNKKNRTKKKPTPKPPVUDE 87
DB 121 ENOESSSSSSSSSTIWKSKNSAANRELQKLVKDNKKNRKKKPTPKPPVUDE 180
QY 88 AGSGLDNGDFKVTPTDSTTQHNKVSTSPKITTAKPINRPSLPNDSKETSITVNYE 147
DB 181 AGSGLDNGDFKVTPTDSTTQHNKVSTSPKITTAKPINRPSLPNDSKETSITVNYE 240
QY 148 TTIVETKETTNTKSTGDEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKG 207
DB 241 TTIVETKETTNTKSTGDEKTTSAKETQSIKTSKADLAPTSKVLAKPTPKAETTTKG 300
QY 208 PALATTPKEPTTPPEKPASTTPKEPTPTTIKSAPTTPKEAPTTTKSAPTTPKPAPTT 267
DB 301 PALATTPKEPTTPPEKPASTTPKEPTPTTIKSAPTTPKEAPTTTKSAPTTPKPAPTT 360
QY 268 KEPAPTTKEAPTTKEAPTTKSAPTTPKEAPTTPKPAPTTPKEAPTTKEAPTTKEPT 327
DB 361 KEPAPTTKEAPTTKEAPTTKSAPTTPKEAPTTPKPAPTTPKEAPTTKEAPTTKEPT 420

```

## RESULT 9

US-10-124-557-48

; Sequence 48, Application US/10124557

; Publication No. US20020137894A1

; GENERAL INFORMATION:

; APPLICANT: Turner, Katherine

; Clark, Stephen C.

; Jacobs, Kenneth

; Hewick, Rodney M.

; Gesner, Thomas G.

; TITLE OF INVENTION: Megakaryocyte Stimulating Factors

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: Massachusetts

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

QY 328 TTPKEPAPTTKEPAPTTKEPAPTTAPKKEPAPTTTPKEPAPTTTPKEPAPTTTKGESPPTPK 387
DB 421 TTPKEPAPTTKEPAPTTKEPAPTTAPKKEPAPTTTPKEPAPTTTPKEPAPTTTKGESPPTPK 480
QY 388 PAPTTPKAPTTTKGAPATTTKSAATTPKEPSPTTTKEPAPTTTPKEPAPTTTPKKEPAPTT 447
DB 481 PAPTTPKAPTTTKGAPATTTKSAATTPKEPSPTTTKEPAPTTTPKEPAPTTTPKKEPAPTT 540
QY 448 KEPAPTTKEPAPTTTKKAPATAPKEPAPTTTPKEPAPTTTPKKEPAPTTTPKKEPAPTT 507
DB 541 KEPAPTTTPKAPATTTTKKAPATAPKEPAPTTTPKEPAPTTTPKKEPAPTTTPKKEPAPTT 600
QY 508 PTPPELAPTTPEEPTTPPEEPAPTTPKAAANPTPKKEPAPTTTPKEPAPTTTPKEPAPTT 567
DB 601 PTPPELAPTTPEEPTTPPEEPAPTTPKAAANPTPKKEPAPTTTPKEPAPTTTPKKEPAPTT 660
QY 569 KETAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATTPKGTAPTT 627
DB 661 KETAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTSKAPATTPKGTAPTT 720
QY 628 PKEPAPTTKEPAPTTPKGTAPTTLKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 687
DB 721 PKEPAPTTKEPAPTTPKGTAPTTLKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 780
QY 688 TPKETAPTTKEPAPTTTPKKEPAPTTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELS 747
DB 781 TPKETAPTTKEPAPTTTPKKEPAPTTTPETPTTSEVSTPTTKEPTTIHKSPDESTPELS 840
QY 748 AEPTPKALENSPKGVPPTTKPAATKPEMTTAKDCTTERDLRTTPETTTAAAPKMTKET 807
DB 841 AEPTPKALENSPKGVPPTTKPAATKPEMTTAKDCTTERDLRTTPETTTAAAPKMTKET 900
QY 808 ATTEKTTESKITATTQVTSITTQDTPPKITLKTTLAPKVTTTKKITTTEIMNKP 867
DB 901 ATTEKTTESKITATTQVTSITTQDTPPKITLKTTLAPKVTTTKKITTTEIMNKP 960
QY 868 EETAKPKDRATNSKATTPKPKQKTKAPKPTSTTKKPKATMPVRKPKTTTPRKMSTMP 927
DB 961 EETAKPKDRATNSKATTPKPKQKTKAPKPTSTTKKPKATMPVRKPKTTTPRKMSTMP 1020
QY 928 LNPTSIKAEAMLTQTTTRPNQNTSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPD 987
DB 1021 LNPTSIKAEAMLTQTTTRPNQNTSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPD 1080
QY 988 MDYLPVFNQGIINPMLS 1006
DB 1081 MDYLPVFNQGIINPMLS 1099

```



; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1361 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:  
US-10-124-557-40  
  
Query Match 99.5%; Score 5343.9; DB 13; Length 1361;  
Best Local Similarity 91.7%; Pred. No. 1.4e-133;  
Matches 1006; Conservative 0; Mismatches 0; Indels 91; Gaps 2;  
  
QY 1 MAWKTLPIVILLLLSVFVIQVSSQ----- 25  
DB 1 MAWKTLPIVILLLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCQYHMECCPDF 60  
  
QY 26 -----ELSCGRCFFSFERGRCDCDAQCKYDKCCPDYESFCA----- 64  
DB 61 KRVCIAELSCGRCFFSFERGRCDCDAQCKYDKCCPDYESFCAEVHNPSTPPSSKAP 120  
  
QY 65 -----EVKONKVRTKKKPKPKPPVVDEAG 89  
DB 121 PPSGASQTKSTKRSPPKPKKKKKVIESEBIETVKONKVRTKKKPKPPVVDEAG 180  
  
QY 90 SGLDNGDFKVTTPDTSTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSVANKET 149  
DB 181 SGLDNGDFKVTTPDTSTQHNKYSTSPKITTAKPINRPSLPNSDTSKETSLSVANKET 240  
  
QY 150 VETKETTTNKQSTDKGKTTSAKETQSTSEKTSADKLAPTSKULAKPTPKAETTKGPA 209  
DB 241 VETKETTTNKQSTDKGKTTSAKETQSTSEKTSADKLAPTSKULAKPTPKAETTKGPA 300  
  
QY 210 LTTPKETPTPKPEASTPKPEPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTKE 269  
DB 301 LTTPKETPTPKPEASTPKPEPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTKE 360  
  
QY 270 PAPTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPTPT 329  
DB 361 PAPTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPTPT 420  
  
QY 330 PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA 389  
DB 421 PKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPAPTTPKEPA 480  
  
QY 390 PTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKE 449  
DB 481 PTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKE 540  
  
QY 450 PAPTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPT 509  
DB 541 PAPTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPT 600  
  
QY 510 TPELAPTTPEPTPTTPEEPAPTPPKAAAPNTPEEPAPTPPKAAAPNTPEEPAPTPPK 569  
DB 601 TPELAPTTPEPTPTTPEEPAPTPPKAAAPNTPEEPAPTPPKAAAPNTPEEPAPTPPK 660  
  
QY 570 TAPTPKGTATTLKPAPTTPKPAKPAKELAPTTTKEPTSTSDKPAPTTPKGTATTPK 629  
DB 661 TAPTPKGTATTLKPAPTTPKPAKPAKELAPTTTKEPTSTSDKPAPTTPKGTATTPK 720  
  
QY 630 EPAPTPKEPAPTTPKGTATTLKPAPTTPKPAKPAKELAPTTTKEPTSTSDKPAPTTP 689  
DB 721 EPAPTPKEPAPTTPKGTATTLKPAPTTPKPAKPAKELAPTTTKEPTSTSDKPAPTTP 780

RESULT 11

US-10-124-557-104  
; Sequence 104, Application US/10124557  
; Publication No. US20020137894A1  
; GENERAL INFORMATION:  
; APPLICANT: Turner, Katherine  
; Clark, Stephen C.  
; Jacobs, Kenneth  
; Hewick, Rodney M.  
; Gesner, Thomas G.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/124,557  
FILING DATE: 16-Apr-2002  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 104:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1140 amino acids







```

; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 876-1170
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:
;
; US-10-124-557-62
;
; Query Match 99.4%; Score 5339.6; DB 13; Length 1404;
; Best Local Similarity 88.2%; Pred. No. 1.8e-133;
; Matches 1006; Conservative 0; Mismatches 0; Indels 134; Gaps 2;
;
; QY 1 MAWKTLPIYLLLSVFVIQVSSQ----- 25
; Db 1 MAWKTLPIYLLLSVFVIQVSSQLSSCAGRGEGYSRDATCNCYDNCQHYMECCPDF 60
;
; QY 26 -----ELSCKGRCFESFERGECDDAQCKKYDKCCPDYESFCAE----- 65
; Db 61 KRVCATLSCKGRCFESFERGECDDAQCKKYDKCCPDYESFCAEVHNFTSPSSKAP 120
;
; QY 66 ----- 65
; Db 121 PPSGASQTIKSTTKRSPKPPNKKTKKVIREEITEEHSVSENQESSSSSSSSSTIW 180
;
; QY 66 -----VKDNKKNRTKKKPTKPPVVDDEAGSLDNGDFKVTTPDTST 106
; Db 181 KIKSSKNSAANRELQKKLVKDNKKNRTKKKPTKPPVVDDEAGSLDNGDFKVTTPDTST 240
;
; QY 107 TQHNKVTSPKITTAKPINRPSLPNPSDTSKETSITVKNKETTIVETKETTITNKQSTDG 166
; Db 241 TQHNKVTSPKITTAKPINRPSLPNPSDTSKETSITVKNKETTIVETKETTITNKQSTDG 300
;
; QY 167 KEKTTSAKETQSIEKTSKDLAPTSKVLAKEPTPKAETTKGPAITTPKEPTTPPKPAS 226
; Db 301 KEKTTSAKETQSIEKTSKDLAPTSKVLAKEPTPKAETTKGPAITTPKEPTTPPKPAS 360

```

```

QY 227 TTPKEPTPTTIKSAPTTTPKEPAPTTTKGAPTTTPKEPAPTTTKGAPTTTPKEPAPTTTPKEPAPTTTKKP 286
;
; Db 361 TTPKEPTPTTIKSAPTTTPKEPAPTTTKGAPTTTPKEPAPTTTKGAPTTTPKEPAPTTTPKEPAPTTTKKP 420
;
; QY 287 APTTTKSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 346
;
; Db 421 APTTTKSAPTTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK 480
;
; QY 347 EPAPAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 406
;
; Db 481 EPAPAPKAPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 540
;
; QY 407 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAP 466
;
; Db 541 TTKSAPTTTPKEPAPTTTPKEPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAPAPTTTPKAP 600
;
; QY 467 APTAPKEPAPTTTPKETAPTTPKLTPTTPKEKAPAPTTTPKEKAPAPTTTPKEKAPAPTTTPKEKAP 526
;
; Db 601 APTAPKEPAPTTTPKETAPTTPKLTPTTPKEKAPAPTTTPKEKAPAPTTTPKEKAPAPTTTPKEKAP 660
;
; QY 527 PEEAPAPTTPKAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTTKGAPTTTKKP 586
;
; Db 661 PEEAPAPTTPKAAPNTPKAPAPTTTPKEPAPTTTPKEPAPTTTPKETAPTTPKGTAPTTTKGAPTTTKKP 720
;
; QY 587 APTTPKAPAPKELAPTTTKETPTSTSDKAPAPTTTPKGTAPTTTPKAPAPTTTPKAPAPTTTPKAP 646
;
; Db 721 APTTPKAPAPKELAPTTTKETPTSTSDKAPAPTTTPKGTAPTTTPKAPAPTTTPKAPAPTTTPKAP 780
;
; QY 647 TAPTTLKAPAPTTTPKAPAPKELAPTTTKGFTSTSDKAPAPTTTPKETAPTTPKETAPTTPKETAPTTPK 706
;
; Db 781 TAPTTLKAPAPTTTPKAPAPKELAPTTTKGFTSTSDKAPAPTTTPKETAPTTPKETAPTTPKETAPTTPK 840
;
; QY 707 KAPAPTTPEPAPTTSEVSTPTTKBPTTIHKSPDESTPELSAPPTPKALENSPKGCVPT 766
;
; Db 841 KAPAPTTPEPAPTTSEVSTPTTKBPTTIHKSPDESTPELSAPPTPKALENSPKGCVPT 900
;
; QY 767 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETAATTTKTTESKITATTTQV 826
;
; Db 901 TKTPAATKPEMTTTAKDKTTERDLRTTPETTTAAPKMTKETAATTTKTTESKITATTTQV 960
;
; QY 827 TSTTTQDTPPFKITTILKTTLAPKVTTTKKTIITTEIMNKPEETAKPKORATNSKATTPK 886
;
; Db 961 TSTTTQDTPPFKITTILKTTLAPKVTTTKKTIITTEIMNKPEETAKPKORATNSKATTPK 1020
;
; QY 887 POKPTKAPKPTSTKPKTMPVRVKPTTPPKMTSTMPELNPTSRIAEAMLTQTTTRPN 946
;
; Db 1021 POKPTKAPKPTSTKPKTMPVRVKPTTPPKMTSTMPELNPTSRIAEAMLTQTTTRPN 1080
;
; QY 947 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLRVVPNQGIINPMLS 1006
;
; Db 1081 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMDYLRVVPNQGIINPMLS 1140

```

Search completed: October 13, 2004, 11:53:08  
Job time : 112.349 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 110.61 Seconds  
(without alignments)  
5233.063 Million cell updates/sec

Title: SEQ1-E  
Perfect score: 5373  
Sequence: 1 MAWKTLPIYLLLSVFIQ.....DMDYLPRVFNQGIINPMLS 1006

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_02: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5339.6	99.4	1404	2 Q92954	Q92954 homo sapien
2	5330.6	99.2	1404	2 Q9BX49	Q9BX49 homo sapien
3	2949.7	54.9	933	2 Q62M25	Q62M25 homo sapien
4	2949.7	54.9	933	2 BAD18580	BAD18580 h cdui fl
5	2307.7	42.9	1054	2 Q9JN99	Q9JN99 mus musculus
6	1335.3	24.9	5179	1 MUC2_HUMAN	MUC2_HUMAN
7	1306.5	24.3	1225	2 Q9VR49	Q9VR49 homo sapien
8	1299.5	24.2	1761	2 Q7KTF6	Q7KTF6 drosophila
9	1299.5	24.2	1761	2 AAS64673	AAS64673 drosophila
10	1227.8	22.9	3150	2 Q7PMD5	Q7PMD5 anopheles g
11	1141.9	21.3	3409	2 Q6SSE6	Q6SSE6 chlamydomon
12	1141.9	21.3	3409	2 AAS07044	AAS07044 chlamydomon
13	1108.6	20.6	1664	1 SLP1_CLOUTM	SLP1_CLOUTM
14	1102.5	20.5	1349	2 Q8WQ4	Q8WQ4 homo sapien
15	1082.4	20.1	3432	2 Q8IR51	Q8IR51 drosophila
16	1082.4	20.1	3458	2 Q8IR52	Q8IR52 drosophila
17	1069.7	19.9	3889	2 Q6SSE8	Q6SSE8 chlamydomon
18	1069.7	19.9	3889	2 AAS07042	AAS07042 chlamydomon
19	1041.7	19.4	1795	2 Q76894	Q76894 drosophila
20	1038.2	19.3	1079	2 Q9N4S7	Q9N4S7 caenorhabdi
21	1023.4	19.0	972	2 Q7OKV7	Q7OKV7 anopheles g
22	1014.7	18.9	1607	2 Q8H6Q5	Q8H6Q5 phytophthor
23	1014.7	18.9	1607	2 AAP74661	AAP74661 phytophthor
24	1008.8	18.8	2284	2 Q7KPK7	Q7KPK7 caenorhabdi
25	1002.8	18.7	1274	2 Q20007	Q20007 drosophila
26	998.5	18.6	9234	2 Q7KTP5	Q7KTP5 drosophila
27	998.5	18.6	9234	2 AAN10531	AAN10531 drosophila
28	990	18.4	1489	2 Q9F449	Q9F449 phytophthor
29	977	18.2	23015	2 Q8IQ18	Q8IQ18 drosophila
30	977	18.2	23015	2 AAN10358	AAN10358 drosophila
31	974.3	18.1	34350	2 Q8WZ42	Q8WZ42 homo sapien

## RESULT 1

ID	Q92954	PRELIMINARY;	PRT; 1404 AA.
AC	Q92954;		
DT	01-FEB-1997 (TrEMBLrel. 02, Created)		
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Megakaryocyte stimulating factor.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]_TaxID=9606;		
RP	SEQUENCE FROM N.A.		
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,		
RA	Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,		
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,		
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;		
RT	"Purification, Biochemical Characterization, and Cloning of a Novel		
RT	Megakaryocyte Stimulating Factor that has Megakaryocyte Colony		
RT	Stimulating Activity.";		
RL	Blood 78:279-279(1991).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Merberg D.M., Fitz L.J., Temple P., Giannotti J., Murtha P.,		
RA	Fitzgerald M., Scaltreto J., Kelleher K., Preissner K., Kriz R.,		
RA	Jacobs K., Turner K.;		
RT	"A Comparison of Vitronectin and Megakaryocyte Stimulating Factor.";		
RL	(In) Preissner K.T., Rosenblatt S., Kost C., Wegerhoff J.,		
RL	Mosher D.F. (eds.);		
RL	BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS., pp.45-52, Elsevier		
RL	Science Publishers B.V. (1993).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	Turner K.J., Fitz L.J., Temple P., Jacobs K., Larson D., Leary A.C.,		
RA	Kelleher K., Giannotti J., Calvetti J., Fitzgerald M., Kriz M.J.,		
RA	Ferenz C., Grobholz J., Fraser H., Bean K., Norton C.R., Gesner T.,		
RA	Bhatia S., Kriz R., Hewick R., Clark S.C.;		
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; U70136; AAB09089.1; -.		
DR	HSSP; P04004; 10C0.		
DR	Genew; HGNC:9364; PRG4.		
DR	GO; GO:0008283; P:cell proliferation; TAS.		
DR	InterPro; IPR000585; Hemopexin.		
DR	InterPro; IPR001212; Somatomedin_B.		
DR	Pfam; PF00045; Hemopexin; 2.		
DR	Pfam; PF01033; Somatomedin_B; 2.		
DR	PRINTS; PR00022; SOMATOMEDIN_B.		
DR	SMART; SM00120; HX; 2.		
DR	SMART; SM00201; SO; 2.		
DR	PROSITE; PS00024; HEMOPEXIN; UNKNOWN 1.		
DR	PROSITE; PS00524; SOMATOMEDIN_B; 2.		
DR	SEQUENCE 1404 AA; 151090 MW; AABD7AD19B35F4F6 CRC64;		

32	973.9	18.1	7962	2	Q10465
33	969.9	18.1	2187	2	P70670
34	966.3	18.0	10578	2	Q8ISF5
35	964.3	17.9	18519	2	Q8ISF6
36	964.3	17.9	18534	2	Q8ISF7
37	960.1	17.9	1458	2	Q757N5
38	960.1	17.9	1458	2	AAS52662
39	956.9	17.8	1121	2	Q7Z884
40	955.4	17.8	5703	1	MUSB_HUMAN
41	955.1	17.8	1480	2	Q9HIE8
42	948.6	17.7	2112	2	Q9VEL9
43	946.8	17.6	926	2	Q9VVG2
44	941.6	17.5	1720	2	Q8I486
45	922.2	17.2	2042	2	Q767L8

## ALIGNMENTS

Query Match		99.4%;	Score 5339.6;	DB 2;	Length 1404;
Best Local Similarity		88.2%;	Pred. No. 2e-97;	0;	Mismatches 134; Gaps 2;
Matches 1006;		Conservative	0;		
QY	1	MAWKTLPIYLLLLSVFVIQQVSSQ	-----	25	
DB	1	MAWKTLPIYLLLLSVFVIQQVSSQ	-----	25	
QY	26	-----ELSKGRCFESFERGECDCDAQCKYDKCCPDYESFCAP	-----	65	
DB	61	KRVCTAELSKGRCFESFERGECDCDAQCKYDKCCPDYESFCAP	-----	65	
QY	66	-----	-----	65	
DB	121	PPSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSNQESSSSSSSSSTIR	-----	180	
QY	66	-----VKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST	-----	106	
DB	181	KIKSSKNSAANRELQKKLVKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST	-----	240	
QY	647	TAPPTLKAPAPKELAPPTTKGTSITSDKPAPTTPKGTAPPTPKAPAPPTPKG	-----	706	
DB	781	TAPPTLKAPAPKELAPPTTKGTSITSDKPAPTTPKGTAPPTPKAPAPPTPKG	-----	840	
QY	707	KPAPPTTETPPPTTSVSTPTTTPKPTTIHKSDESTPELSAPPTPKALENSPKPGVPT	-----	766	
DB	841	KPAPPTTETPPPTTSVSTPTTTPKPTTIHKSDESTPELSAPPTPKALENSPKPGVPT	-----	900	
QY	767	TKTPAATKPBMTTAAKDKTERDRLTTPETTTAAKPMKETAATTEKTSKITAATTTQV	-----	826	
DB	901	TKTPAATKPBMTTAAKDKTERDRLTTPETTTAAKPMKETAATTEKTSKITAATTTQV	-----	960	
QY	827	TSITTTQDTTTPKTIITLLKTIILAPKVTITTKTIITTEIMNKPEETAKPKDRATNSKATTPK	-----	886	
DB	961	TSITTTQDTTTPKTIITLLKTIILAPKVTITTKTIITTEIMNKPEETAKPKDRATNSKATTPK	-----	1020	
QY	887	PQKFTKAPKPTSTTKPKTMRVRKPKTTPTRKMTSTMPNLNPTSRIAEAMLQTTTRPN	-----	946	
Query Match		99.2%;	Score 5330.6;	DB 2;	Length 1404;
Best Local Similarity		88.1%;	Pred. No. 3e-97;	2;	Indels 134; Gaps 2;
Matches 1004;		Conservative	0;	Mismatches	
QY	1	MAWKTLPIYLLLLSVFVIQQVSSQ	-----	25	
DB	1	MAWKTLPIYLLLLSVFVIQQVSSQ	-----	25	
QY	26	-----ELSKGRCFESFERGECDCDAQCKYDKCCPDYESFCAP	-----	65	
DB	61	KRVCTAELSKGRCFESFERGECDCDAQCKYDKCCPDYESFCAP	-----	65	
QY	66	-----	-----	65	
DB	121	PPSGASQTIKSTTKRSPKPNKKTKVIESEBITTEHSVSNQESSSSSSSSSTIR	-----	180	
QY	66	-----VKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST	-----	106	
DB	181	KIKSSKNSAANRELQKKLVKONKNRTKKKTPKPPVVDVDEAGSLDNGDFKVTTPDTST	-----	240	
QY	107	TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTSDG	-----	166	
DB	241	TQHNKYSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTITNKQSTSDG	-----	300	
QY	167	KEKTTSAKETQSTKETSADKLAFTSKVLAKPTPKAETTTKGPALTTPKESPTTPPKEPAS	-----	226	
DB	301	KEKTTSAKETQSTKETSADKLAFTSKVLAKPTPKAETTTKGPALTTPKESPTTPPKEPAS	-----	360	
QY	227	TTPKEPTPTTIKSAPTTPKAPPTTTTKSAPTTTPKAPPTTPKAPPTTPKAPPTTTKBP	-----	286	
DB	361	TTPKEPTPTTIKSAPTTPKAPPTTTTKSAPTTTPKAPPTTPKAPPTTPKAPPTTTKBP	-----	420	
QY	287	APTITTKSAPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPK	-----	346	
DB	421	APTITTKSAPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPKAPPTTPK	-----	480	



```

Db 471 ATTTEKTTESKITTATTQTTSITTQDTTTFKLTLLTKTTLAPKVTITTKITITTEIMNKP 530
QY 868 EETAKPKDRATNSKATTPKPKQPTKAPKKTSTKPKKTMKPRVRKPKTTPPKMTSTMP 927
Db 531 EETAKPKDRATNSKATTPKPKQPTKAPKKTSTKPKKTMKPRVRKPKTTPPKMTSTMP 590
QY 928 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 987
Db 591 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 650
QY 988 MDYLPVNPQGIINPMLS 1006
Db 651 MDYLPVNPQGIINPMLS 669

RESULT 4
BAD18580 PRELIMINARY; PRT; 933 AA.
AC BAD18580;
DT 12-MAY-2004 (TrEMBLrel. 27, Created)
DT 12-MAY-2004 (TrEMBLrel. 27, Last sequence update)
DE CDNA FLJ16561 fis, clone SYN0V4003981, moderately similar to Homo
DE sapiens proteoglycan 4, (megakaryocyte stimulating factor, articular
DE superficial zone protein, camptodactyly, arthropathy, coxa vara,
DE pericarditis syndrome) (PRG4).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TTSUE=Synovial membrane tissue;
RA Ota T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma A., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuko Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK131434; BAD18580.1; -.
SQ SEQUENCE 933 AA; 102512 MW; B883773C5BC21A23 CRC64;

Query Match 54.9%; Score 2949.7; DB 2; Length 933;
Best local similarity 52.4%; Pred. No. 1e-50;
Matches 576; Conservative 0; Mismatches 0; Indels 523; Gaps 2;

QY 1 MAWKTPYILLLLSVFVIOQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
Db 1 MAWKTPYILLLLSVFVIOQVSSQELSCKGRCFESFERGECDCDAQCKYDKCCPDYE 60
QY 61 SPCAEVHNTPSPSSKXKAPPSPGASQIKSTTKSPKPNKXKTKKVIIEEITEHSVS 120
Db 61 SPCAEVHNTPSPSSKXKAPPSPGASQIKSTTKSPKPNKXKTKKVIIEEITEHSVS 120
QY 66 -----VKDNKNRTKKKXPTPKPPVDE 87
Db 121 ENQESSSSSSSSSSSTIRKIKSSKNSAANRELQKLKVKDNKNRTKKKXPTPKPPVDE 180
QY 88 AGSLDNGDFKVTTPDSTTQHNKVSIPKITTAKPINPRLSPNPSDTSKETSLTWNKE 147
Db 181 AGSLDNGDFKVTTPDSTTQHNKVSIPKITTAKPINPRLSPNPSDTSKETSLTWNKE 240
QY 148 TTVEKTKTTNKTSTDGKEKTTSAKETSIETSAKDLAPTSKVLAKPTPKAETTTKG 207
Db 241 TTVEKTKTTNKTSTDGKEKTTSAKETSIETSAKDLAPTSKVLAKPTPKAETTTKG 300
QY 208 PALTPKPEPTTPPKPEASTTPKBPPTTTKSAPTTPKPEAPTTKSGAPTTPKBPPTTT 267
Db 301 PALTPKPEPTTPPKPEASTTPKBPPTTTKSAPTTPKPEAPTTKSGAPTTPKBPPTTT 360

```

```

QY 268 KEPAPTTPKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTTPKKAPATTTPKEPAPTTTPKEPTP 327
Db 361 KEPAPTTPKE----- 370
QY 328 TTTPKEPAPTTTKEPAPTTTPKEPAPTTAPKKAPATTTPKEPAPTTTPKEPAPTTTKGSPSTTPKE 387
Db 371 ----- 370
QY 388 PAPTTPKSAPTTTKEPAPTTTKEPAPTTTKSAPTTTPKEPAPTTTKEPAPTTTPKEPAPTTTPKAPATTTP 447
Db 371 ----- 370
QY 448 KEPAPTTPKEPAPTTTKEPAPTTAPKKAPATTTPKEPAPTTTPKKTTPTPKLAAPTTPKEPA 507
Db 371 ----- 370
QY 508 PTTPELAPTTPEEPTTPTEBPAPTTPKAAAPNTPKBPAPTTTPKEPAPTTTPKEPAPTTTP 567
Db 371 ----- 370
QY 568 KETAPTTPKGTAPTTILKEPAPTTTPKPKAPKELAPTTTKEPTSTTSKAPATTTPKGTAPTT 627
Db 371 ----- 370
QY 628 PKEPAPTTTPKEPAPTTTPKGTAPTTTLKEPAPTTTPKPKAPKELAPTTTKEPTSTTSKAPATTTP 687
Db 371 ----- 370
QY 688 TPKETAPTTTPKEPAPTTTPKPKAPATTTPETPPPTTSVSTPTTKEPTTIHKSPDESTPELS 747
Db 371 -----PAPTTPEPTPTTSVSTPTTKEPTTIHKSPDESTPELS 410
QY 748 AEPTPKALENSPKERGVPPTTKTAAATKPEMTTTAKDKTTERDLRTTPETTTAAAPKMTKET 807
Db 411 AEPTPKALENSPKERGVPPTTKTAAATKPEMTTTAKDKTTERDLRTTPETTTAAAPKMTKET 470
QY 808 ATTTEKTTESKITTATTQTTSITTQDTTTFKLTLLTKTTLAPKVTITTKITITTEIMNKP 867
Db 471 ATTTEKTTESKITTATTQTTSITTQDTTTFKLTLLTKTTLAPKVTITTKITITTEIMNKP 530
QY 868 EETAKPKDRATNSKATTPKPKQPTKAPKKTSTKPKKTMKPRVRKPKTTPPKMTSTMP 927
Db 531 EETAKPKDRATNSKATTPKPKQPTKAPKKTSTKPKKTMKPRVRKPKTTPPKMTSTMP 590
QY 928 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 987
Db 591 LNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 650
QY 988 MDYLPVNPQGIINPMLS 1006
Db 651 MDYLPVNPQGIINPMLS 669

RESULT 5
Q9JW99 PRELIMINARY; PRT; 1054 AA.
AC Q9JW99;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus.
DE Name=Pr94;
DE Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20573856; PubMed=11124536;
RA Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT "Isolation, characterization and mapping of the mouse and human PRG4
RT (proteoglycan 4) genes.";

```



DISULFID	5102	5154	By similarity.
DIFULFID	?	5159	By similarity.
CARBOHYD	163	163	N-linked (GlcNAc. .) (Potential).
CARBOHYD	423	423	N-linked (GlcNAc. .) (Potential).
CARBOHYD	670	670	N-linked (GlcNAc. .) (Potential).
CARBOHYD	770	770	N-linked (GlcNAc. .) (Potential).
CARBOHYD	894	894	N-linked (GlcNAc. .) (Potential).
CARBOHYD	1139	1139	N-linked (GlcNAc. .) (Potential).
CARBOHYD	1154	1154	N-linked (GlcNAc. .) (Potential).
CARBOHYD	1215	1215	N-linked (GlcNAc. .) (Potential).
CARBOHYD	1230	1230	N-linked (GlcNAc. .) (Potential).
CARBOHYD	1246	1246	N-linked (GlcNAc. .) (Potential).
CARBOHYD	1787	1787	N-linked (GlcNAc. .) (Potential).
CARBOHYD	1820	1820	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4339	4339	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4351	4351	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4362	4362	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4373	4373	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4422	4422	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4438	4438	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4502	4502	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4616	4616	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4627	4627	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4752	4752	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4787	4787	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4881	4881	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4888	4888	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4955	4955	N-linked (GlcNAc. .) (Potential).
CARBOHYD	4970	4970	N-linked (GlcNAc. .) (Potential).
CARBOHYD	5019	5019	N-linked (GlcNAc. .) (Potential).
CARBOHYD	5038	5038	N-linked (GlcNAc. .) (Potential).
CARBOHYD	5069	5069	N-linked (GlcNAc. .) (Potential).
CONFLICT	1351	1351	H -> L (in Ref. 3).
CONFLICT	1412	1412	T -> S (in Ref. 3).
CONFLICT	1449	1449	L -> P (in Ref. 3).
CONFLICT	1504	1504	M -> T (in Ref. 3).
CONFLICT	4192	4192	G -> S (in Ref. 2).
SEQUENCE	5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;		
Query Match	24.9%; Score 1335.3; DB 1; Length 5179;		
Best Local Similarity	24.3%; Pred. No. 3,9e-18;		
Matches	418; Conservative	65; Mismatches 459; Indels 777; Gaps 63;	
2	AWKTLPIY-----LLLLSVFVI-----QQVSSQE 26		
1020	SWKEAFTCPDVSTNPEPCSINPHRRSWAEKQCSILKSSVSFCSHKVDPKPFYEAQVHDS 1079		
27	LSCK--GRCPESPERGECDDA-----QCKYDKC-----CP---DY-----59		
1080	CSDTGDGC-----ECFSAVASAQECTKEGACVFWRTPLDCLPCFDYVNPPECE 1131		
60	-----ESF--CAEV-----KONKQRTKK-----77		
1132	WHYPCNRRSFETCRTINGIHNSISVSLGCVPRCPKDRPIYEDLKKCVTADKCGYV 1191		
78	-PTKPP-----VDEAGSGL-----DNG-----95		
1192	EDTHYPGASVPTEETCKSCVCTNSSQVCRPEBGKILNTQDGAFCYWEICGNGIVEK 1251		
96	-----DFKVTTPDSTTTOHNKVSFSPK-----117		
1252	HFNICSITTRPSTLTFTTITLPTTPTSFTTTTTTTTPTSGTSLSTTPKLCLMSDWINE 1311		
118	-----117		
1312	DHPSSGDDGDRPEFDGVCGAPEDIICRSVKDPHLSLEQHGQKQVQDVSVGFCKNEDQF 1371		
118	-----IITAKPINRPSLPPNSDTSKETSLTWNKETTVEIK 153		
1372	GNGPFGLCYDKIRVNCWPMDCIITPESPPTTTPSPPTTTTTLPTTTPSPPT-----1427		
154	ETTTNKTQSTDGKEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTKGPAITTP 213		







OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 [1]  
 RA SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.P.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champagne M., Pfeiffer B.D.,  
 RA April J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrieli A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong M., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195 (2000).  
 [2]  
 RA SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
 RT *melanogaster* euchromatic genome sequence."  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).  
 [3]  
 RA SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomics perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).  
 [4]  
 RA SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Froehner S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,

RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review."  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).  
 [5]  
 RA SEQUENCE FROM N.A.  
 RX FLYBASE;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 [6]  
 RA SEQUENCE FROM N.A.  
 RX FLYBASE;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR ENBL: AE003626; AAS64673.1;  
 DR InterPro: IPR011009; Kinase\_like.  
 FT NON TER 1  
 SQ SEQUENCE 1761 AA; 197521 MW; 62A5E16E6241B3F4 CRC64;  
 Query Match 24.2%; Score 1299.5; DB 2; Length 1761;  
 Best Local Similarity 24.2%; Pred. No. 5.5e-18;  
 Matches 441; Conservative 114; Mismatches 361; Indels 905; Gaps 82;  
 QY 12 LLLSVFV-IQQVSSQBLSCCKGRCFESFERGECDCDAQCKKYDKC-----CPDYESP 62  
 Db 1 VVLSCEIWLALYHSANAALG-----SEIKGDESTTESAEKYDYCWINPLPGCP----- 52  
 QY 63 CAEVKONKNNRTKKKTPKPPVVDGAGLNDGDFKVTTPDTSSTOHNVSTSPKITT-- 120  
 Db 53 -----FGPGDGNSTTKISTLATTKSTVTITSEETLTK 85  
 QY 121 -----AKPINRPSLPNSDTSKETSITVAKETIVETKNTKQSTGKE-----K 169  
 Db 86 ITTIKSTAKPTQKTINPTTE-----KITTPKATII--KSTATTARATEAPKTEQTLR 138  
 QY 170 TTSAKETQ-----SIEKTSADLAPTSKVLAKPPKPAETTKGPAITTPKEPT-- 217  
 Db 139 TTIKSTSELIITLTKITTKSAETKSKSHNTTKKSTT-----LRITEPTTKKSTAK 193  
 QY 218 -----PTTPKEPASTTPKEPT-----PTTKSA- 245  
 Db 194 TTREPTTKRETERUTQEPSTSKTTHETTABPATKKTTHETPTQKSTTLRITEPTTK 253  
 QY 246 EPAPTTKSAPTTPKEPAPTTPKEPA----- 271  
 Db 254 SSTAKITRE-PTTKGETTERTTKEPSTSKTTHETTABPATKKTTHETPTQKSTTLRITE 312  
 QY 272 -----PTTPKEPAPTTPKEPA----- 287  
 Db 313 EPTTRKSSSTARTRPTTKRETTERTTQEPSTSKTTHETTABPATKKTTHETPTQKSTT 372  
 QY 289 -----PTTKSA-----PTTPKEPAPTTPKKA- 313  
 Db 373 LRITEPTTRKSSAKTTRPTTKRETTERTTKEPSTSKTTHETTABPATKKTTHETPT 432  
 QY 314 PK-----EPA-----PTTPKEPTTPKEPA-----PTTKPA- 341  
 Db 433 QKSTTLRITEPTTRKSSAKTTRPTTKRETTERTTQEPSTSKTTHETTABPATKKT 492  
 QY 342 -----PTTPK-----EP-----APTA 352  
 Db 493 HEPTTQKSTTLRITEPTTRKSSAKTTRPTTKRETTERTTQEPSTSKTTHETTABPAT 552  
 QY 353 PKKA-PTTPK-----EPA-----PTTPKEPAPTTPKEPS----- 381  
 Db 553 TKKTTHETPTQKSTTLRITEPTTRKSSAKTTRPTTKRETTERTTQEPSTSKTTHET 612  
 QY 382 -----PTTPKEPA-----PTTKSA-----PT- 398  
 Db 613 TAPATKKTTHETPTQKSTTLRITEPTTRKSSAKTTRPTTKRETTERTTQEPSTSKT 672  
 QY 399 -----TTKEPA-----PTTKSA-----PTTPKPSPTTPKPEP 426  
 Db 673 TTHTTAPATKKTTHETPTQKSTTLRITEPTTRKSSAKTTRPTTKRETTERTTKEP 732

QY 427 A-----PTTPK-----BPA-----PTTPKKAPT 445  
 DB 733 TTRKTTTHKTTBEPTTKTTTBEPTTKSTLTKTEPTTKRKTSTTKTTEPTTKRKTTER 792  
 QY 446 TPKPA-----PTTPKEPAPTTTKKPA-PTAPK-----EPA-----475  
 DB 793 TTKEPTTKRKTTHKTTBEPTTK-----TTTKTTTBEPTTKSTLTKTEPTTKRKTSTTK 848  
 QY 476 -----PTTPKETA-----PTTPKKLT-----PTT-P 495  
 DB 849 TTREPTTKETTERTTQEPSTSKTTTHTTAEPATKKTTHTEPTTKSTLRLITEPTTKR 908  
 QY 496 EKLAPTTPKPAPTTPEELAPTTPEPT-----PTTPE---EPA-----531  
 DB 909 SSTAKTTRE---PTTKRETTERTTKEPTTKRKTTHKTTBEPTTKKTTHTEPTTKSTLTKP 965  
 QY 532 ---PTTPKAAAPNTPKEPAPTTPEAPTTPKPA-----PTTPKETA-----571  
 DB 966 TTEPTTKRKTSTTKTRE---PTTKRKTERTTKEPTTKRKTTHKTTTEPTTKNTTKTKTH 1023  
 QY 572 -PTTPKGTA-----PTTLKEPAPTTPKKPAKELA-----PTTKGPTSTSTSKPA- 616  
 DB 1024 EPTTKSTLTKPTEPTTKRKTSTTKTTRPTTKRKTTHTEPTTKRKTTHTEPTTKRKT 1083  
 QY 617 -----PTTPKGTA-----PTTPKEPAPTTPKPA-----PTTPKGTAPTTL 652  
 DB 1084 KKTTHTEPTTKSTLRLITEPTTKRKTSTAKTTTKEPTTKRKTTHTEPTTKRKTTHKTT 1143  
 QY 653 KEPA-----PTTPK-----KPAPKELAPTTTKGPTSTSTSKPAPTTPEAPTTPKPA 701  
 DB 1144 EPTTKKTTHTEPTTKSTLTKPTEE---PTTRK---TSTTKTTEPTTKRKTTHTEPTTKRKT 1198  
 QY 702 -----PTTPK-----KP-----APTPEPTPPPTSEV 723  
 DB 1199 TRKTTTHKTTTEPTTKNTTKTTTBEPTTKSTLTKPTEPTTKRKTSTTKTTEPTTKRV 1258  
 QY 724 STPTTKPEP-----TTTHKSDE-----STPELSABEPTPKALENSKEPGVPTTKPAAT 773  
 DB 1259 TTEPTTKRKTTHKTTTEPTTKKTTHTEPTTKKTTHTEPTTKK-STLTKPTEPTTKRKTST 1317  
 QY 774 KPENMTTAKDKTTERDLR-----KTITTTIIMNKP-----TTPETTTAAP-- 801  
 DB 1318 KTTRPTTKRKTTHTEPTTKRKTTHKTTTBEPTTKKTTHKTTTBEPTTKKSTLTKPTE 1377  
 QY 802 -----KMTKEPAT---TTEKTESKITATTTQVSTTTQDTPPTPKITLTKT---844  
 DB 1378 EPTTKRKTSTTKTTRPTTKRKTTHTEPTTKRKTTHTEPTTKRKTTHTEPTTKRKTTHTEP 1435  
 QY 845 -----TTLAP-----KVTTTK-----KTITTTIIMNKP-----EETAKPKDRA 877  
 DB 1436 TTKKSTLTKPTEPTTKRKTSTTKTTRPTTKRKTTHTEPTTKRKTTHKTTTTEPTTKK 1495  
 QY 878 TNSKATTPKPO-----KP-----TKAPKKPT-----898  
 DB 1496 TTKKTTHTEPTTKSTLTKPTEPTTKRKTSTTKATREPTTKRKTTHTEPTTKRKTTHTEP 1555  
 QY 899 -----STPKPKTMVRVKPTTTPRKMSTWPELNPSRIEAM 938  
 DB 1556 KTTEPTTKKTTHKTTTBEPTTKSTL-----KTEPTTKRKTSTTKTTRPTTKRKTSTTS 1611  
 QY 939 LQTT-----TRPQNTNSKLVENPKSEDAGGAGET-----970  
 DB 1612 VKTTADTTKRTTAEMSTTQEPSTSVETTTNSQNSQNTTSTTTERQHVHHHHHHIHYH 1671  
 QY 971 -----PHMLLRPH-----978  
 DB 1672 KPADLGSILPL 1731  
 QY 979 ---VEMPEVT-PDMDYLRVFN 996  
 DB 1732 TAILSLPEISLPLPLPLPLPN 1752

RESULT 9  
 AAS64673 PRELIMINARY; PRT; 1761 AA.  
 ID AAS64673;  
 AC AAS64673;  
 DT 01-APR-2004 (TREMBlrel. 27, Created)  
 DT 01-APR-2004 (TREMBlrel. 27, Last sequence update)  
 DT 01-APR-2004 (TREMBlrel. 27, Last annotation update)  
 DE CG33300-PA (Fragment).  
 DE CG33300.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,  
 RA Brandon R.C., Rogers J.H., Blazek E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhattacharya D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mout S.M., Moy M., Murphy B., Murphy L., Murthy D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rainert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 melanogaster euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,



```

Db 1612 VKTTADQTKRTAEMSTTNQBPSTVETTTNSNOSNTTSTTETTBEOHHHHHHHHIYH 1671
QY 971 -----PHMLLRPH----- 978
Db 1672 KEADLGPSILPLDPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPLPL 1731
QY 979 --VFMEVPT-PDMOYLPRVFN 996
Db 1732 TAISLPEISLPLPLPLPLPN 1752

RESULT 10
Q7PMD5 PRELIMINARY; PRT; 3150 AA.
AC Q7PMD5;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000004655 (Fragment).
GN Name=ENSANGG0000003651;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAA01008980; EAA13969.2; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR006770; OGF_recept.
DR Pfam; PF04680; OGF_I,II; 80.
FT NON TER 1
FT NON TER 3150 3150
SQ SEQUENCE 3150 AA; 322879 MW; 3C7B3D441CE8C839 CRC64;

Query Match 22.9%; Score 1227.8; DB 2; Length 3150;
Best Local Similarity 19.9%; Pred. No. 2.8e-16;
Matches 394; Conservative 103; Mismatches 388; Indels 1092; Gaps 72;

QY 74 TKKKPTKPPVDEAGS-----GLDNGDFKVTTPDTSTQHN-- 110
Db 895 TPRPTDTMSSASTPEPSTPGTTRTTPRTSTESTDTMSSASTPEPSTPGTTR 954
QY 111 -----KVSTSPKITTAKINP----- 126
Db 955 TTPTPTSTESTDTMSSASTPEPSTPGTTRTTPRTSTESTDTMSSASTPEPSTKP 1014
QY 127 -----RSLPNSTSKETSLTVNKETTVETKETTNNKQSTDGKKTSA---- 173
Db 1015 GTTRTPTPTSTESTDTMSSASTPEPSTPGTTRTTPRTSTESTDTMSSASTPE 1073
QY 174 -----KETQSIKTSKADLAP-----TSKVLAKPTPKAET----- 203
Db 1074 PSTTPTPTPTPTPTPTSTESTDTMSSASTPEPSTPGTTRTTPRTSTESTDTMSSA 1133
QY 204 -----TKGALUTPKPEPP-----TTPKEPAS----- 226
Db 1134 STPEPSTPGTTRTTPRTPTSTESTDTMSSASTPEPSTPGTTRTTPRTSTESTDTT 1193
QY 227 -----TTPKEP-----TPTIKSAPTT----- 244
Db 1194 MSSASTPEPSTPTPTPTPTPTSTESTDTMSSASTPEPSTPGTTRTTPRTPTSDT 1253
QY 245 -----KEPA--PTTKSAPTT----- 259

```

```
QY 715 -----TPPPTT-----SEVSTPTTTKBPETTHKSP-----739
D 2393 DTTMSSASTPESTTGGTTRTTPRTPTDSTMSASTPESTTGGTTRTTPRTSTES 2452
QY 740 -----DESPBELAEP-----TPKALENSPKEP-762
D 2453 TDTMSSASTPESTTGGTTRTTPRTSTESTDSTMSASTPESTTGGTTRTTPRT 2512
QY 763 ---GVPTTKTAAKPEWTTAKDTERDLRTTPETTTAAKPKMTKETAIT-810
D 2513 STSTDTMTAST-PEPSIT-PDTRTTPRTPTDSTMSASTPESTTGGTTRTTP 2570
QY 811 -----TBKTTESKITATTQVTSITTDTPFKITLLKTTTLA 848
D 2571 RPTPTDSTMSASTPESTTGGTTRTTPRTSTESTDSTMSASTPESTTGGTTRT 2630
QY 849 P-----KVTTTKTITTEIMNKPEBTAKPKDRATN 879
D 2631 PTPPTDSTMSASTPESTTGGTTRTTPRTSTESTDSTMSASTPESTTGGTTRT 2690
QY 880 -----SKATTPKPO-----KPTK-----892
D 2691 PTPPTDSTMSASTPESTTGGTTRTTPRTPTDSTMSASTPESTTGGTTRT 2750
QY 893 APKPTGSK-----KPTMP---RVKPKTTPPRKWTSTM-----PELNP 930
D 2751 TPTPTSTESTDSTMSASTPESTTGGTTRTTPRTPTDSTMSASTPESTTGGTTRT 2810
QY 931 -TSRIAEAMLTQTRPNQTPNSKLVEVPKSEDAGAGETPHMLRPHVFMPEVTP 986
D 2811 GTTR-----TTPTPTDSTMSASTPESTTGGTTRTTPRTPTDSTMSASTPESTTGGTTRT 2843

RESULT 11
Q6SSE6 PRELIMINARY; PRT; 3409 AA.
AC Q6SSE6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Plus agglutinin.
GN Name=SAG1;
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Waifenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY450930; AAS07044.1;
DR InterPro; IPR002951; Atrophin.
DR InterPro; IPR003882; Pibcil_extensin.
DR PRINTS; PRO1222; ATROPHIN.
DR PRINTS; PRO1218; PSTLEXTENSIN.
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 21.3%; Score 1141.9; DB 2; Length 3409;
Best Local Similarity 17.9%; Pred. No. 1.5e-14;
Matches 303; Conservative 127; Mismatches 365; Indels 898; Gaps 47;

QY 76 KKPTKPPVVDAGSLDNGDFKVTTPDT---STQHKNKVSPTTKTAKPINPRESLPP 132
D 574 RAPRESPP-----FHPSPPDPPASSVPPSPKPPPPAPSPAPSP 620
QY 133 NSDTSKETSLTVNKETTVEKTTTTNKQTSIDGKEKTSKQTSKTSKQTSK 192
D 621 -----PSP 624
QY 193 VLAKPTPKAEITTKGPAITTPKEPTPTTKPEAST-----TPKEPTTTIKSAFTTPK 245
D 193 -----TPKEPTTTIKSAFTTPK 245
```

```
Db 625 APPSPAPPS-----PAPSPQPPSPVPPQPPSPKPPSPAPSPSPVPPSPAPSP 678
QY 246 EPAPTTTKSAPTTTKKEPAPTTPK-----EPAPTTPKBPATTTTKSAPTTPKEP 301
D 679 PPSAPENPAPSPAPLPSPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 738
QY 302 APPTTKKPAFT-----TPKEPAPTTPKBPATTTTKKEPAPTTPKBPATTTKEP 347
D 739 APPSPPEPPAPSPPEPPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 798
QY 348 PAPT-----APKCAPPTTPKEPAPTTPKEPAPT-----375
D 799 PSAPSPAPSPPEPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 858
QY 376 -TTKEPSPPTTKBPATTTTKSAPTTTKKEPAPTTPKSPAPTTPKEPAPTTPKEP 434
D 859 PAPSPAPSPPEPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 917
QY 435 AP-----TPKCAPPTTPKEPAPTTPKBPATTTTKBPATTTKBPATTPAK-472
D 918 SPPEPPSPPEPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 977
QY 473 -----EPAPTTPKBPATTTTKKLTPTTPKLPAPTPE---KCAPTTPKE 513
D 978 PPSDPSPAPSPDPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1036
QY 514 LAFT-----TPPEPTTTPKEPAPTTPKAAAPN-----TPKEPAPTTPK 552
D 1037 TPPTSPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 1096
QY 553 EPAPTTPK-----EPAPTTPKBPATTTKBPATTT-----583
D 1097 PPSAPSPPEPPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1156
QY 584 -----KEPAPTTPKBPAPKELAPT-----TKPT 608
D 1157 SPPEVPTTPPPPPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1216
QY 609 STTSKAPATTPKGTAPTTPKEPAPTTPKBPATTTKBPATTTKBPATTTKBPAPKEL 668
D 1217 SPEPPSPAPSPAPSP-APSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1271
QY 669 APT-ITKGTSTTSKBPATTPKEPAPTTPKBPATTTKBPATTTKBPATTTKBPAT 726
D 1272 SPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAP 1328
QY 727 TTTKEPTTTTHKSPDESTPELSAETPKALEN-----SPKEPGV-764
D 1329 AAALPPLPSPAPPLPVPFAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSPAPSP 1388
QY 765 ---PTTKTP-----AATKE-----776
D 1389 PPPPTPTPLAPLPDCTLLAQALLSIPDAANSVVFVSAGLPTSVARSTPELLASFC 1448
QY 777 -----MTTTP-----781
D 1449 VCSCLTATSLVGGSSRGNNNTGSSGGNYISNGGDAAIQRIAGNSTDRGSSSGS 1508
QY 782 -----KDKTTERDL-----RTTP-----794
D 1509 GSSSWGPGTETAEEAWDAVDQGYQLQLSGGYTTRTVVDRTPPVSNGNVTLSANRI 1568
QY 795 -----ETTTAAPKMT--804
D 1569 KQEPSAVGEASINALGSKQAMLLTISFSEVPAPFAPDAASLIVTGALVAEWAAKMTFY 1628
QY 805 -----KETATTTEKITESKIT-----820
D 1629 VLAMTLPALVATAAGSSSSGTSRSGNGNGTAAAAAABPPAGTTGRRRALQQQAAA 1688
QY 821 -----820
D 1689 PPPFASGSSSLSGAANTANQQORHVHLLPATAYADAARNPGRNDLSLSVELTDNAVSP 1748
```

```

QY 821 -----ATTQVSTTTQDTPFKITTLTKTTTLAPKVT----- 852
Db 1749 AVGEALATTARVTAATYP-----AAATTTLVAASSSFAQAIKAGSLIQSYHIQM 1801
QY 853 -----TTKKTITTT----- 862
Db 1802 LTMSLYLASRGVGREYGEYAVFKYAVLGVKGNLGPAAEAMPTNEKEVTAABQARQVGD 1861
QY 863 -----IMNKPEETAKPKRATNSKATTPKQPK----- 890
Db 1862 LMPIGNDLLGSSNTTASGSSSSSSNSP-PRRPPPPAAGSTGLLFSNADASPPPLA 1920
QY 891 -----TKAPKKPTSTKKPKTMVRKPKTTPPRKMTST--MPELN- 929
Db 1921 VATPALPAELPSTIAAAATAAP-----PRLPSP--PPPAVGSGTGLVLRHL 1965
QY 930 -----PSTRAEAMLQTTTRNQTPNSKLYE----- 955
Db 1966 MOQMLQPPAAVAAPPPPPASSALVLQSPPPPPPPSOLLIQOASATVSDMDLLYL 2025
QY 956 -----VNPKGEDAGGAGETPHMLLRPHVFMVEVTPDMDYL 991
Db 2026 VVAAMLITAVAGRLIAAIVLRLVSP-----PHPLA-----F 2060
QY 992 PR----- 993
Db 2061 PRELTTIAGLILVALTFYSCMALGSPAADWHGSRTAAYCVLITIAVVFYAAFLMWLALARA 2120
QY 994 --VPMOGLINPM 1004
Db 2121 WMVP-OFTLVEPM 2132

RESULT 12
AAS07044 PRELIMINARY; PRT; 3409 AA.
AC AAS07044;
DT 02-MAR-2004 (TEMBLrel. 27, Created)
DT 02-MAR-2004 (TEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TEMBLrel. 27, Last annotation update)
DE Plus agglutinin.
GN SAGI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonad.
OC NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferris P.J., Wafenschmidt S., Umen J.G., Ishida K., Kubo T., Lau J.,
RA Goodenough U.W.;
RT "Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY450930; AAS07044.1;
SQ SEQUENCE 3409 AA; 336045 MW; 7FE87633EDD6631F CRC64;

Query Match 21.3%; Score 1141.9; DB 2; Length 3409;
Best Local Similarity 17.9%; Pred. No. 1.5e-14;
Matches 303; Conservative 127; Mismatches 365; Indels 898; Gaps 47;

QY 76 KKPTPKPPVDEAGSLDNGDFKVTTPDT---STQHKNKYSTSPKITTAKPINRPSLPP 132
Db 574 RAPRPSPP-----FHPSPDSPASSVPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 620
QY 133 NSDTSKETSIFVNKETTIVETKETTNNKQTSIDGKETSIAKETQSIKTSAKDLAPTSK 192
Db 621 -----PPSP 624
QY 193 VLAKPTKPAETTKGPAULTTKPEPTTPKPEPAST-----TPKEPTTIKSAPTPK 245
Db 625 APPSPAPPS-----PAPSPQPPSPVPQPPSPVPSPSPSPSPSPSPSPSPSPSPSP 678
QY 246 EPAPTTIKSAPTTKPEAPTTTKPEAPTTK-----EPAPTTTKPEAPTTTKSAPTPK 301

```

```

Db 1749 AVGEALATTARVTAAATYP-----AVAAATTLVAAASSSFAQAIKRAKSLQGSYHIQM 1801
Qy 853 -----TTTKTITITE-----862
Db 1802 LTMSLYLASRGVGEYGEYAVEFKYAVLVKGNLGPAAEAMPTNEKEVTAABEQARQVGGD 1861
Qy 863 -----INMKPEETAKPDRAVNSKATTPKQKP-----890
Db 1862 LWPIGNDLLGSSNTTASGSSGSSSSNSP-PRRPPPPPAAGSTGLLFSNADASPPLA 1920
Qy 891 -----TKAPKPTSTKKEKTPVRKPTTTPPKMTST--MPENL- 929
Db 1921 VATPALPAPLPSTTAAATAAP-----PRLPSP---PPPAVGSSTGVLPRRHL 1965
Qy 930 -----PTSRIBAEAMLQTTTRPNQTPNSKIVE-----955
Db 1966 MQQMLQPPAAVAAPPPPASSSALVLQSPPPPPPSQLLIQASATYVSDMQDLLYTL 2025
Qy 956 -----VNPKSEDAGGAEGETHPMLLRPHVFMPEVTPDMDYL 991
Db 2026 VVAAMLITVAAGRLIAAVALYRLIVSPE-----PPFLA-----F 2060
Qy 992 PR-----993
Db 2061 PRLTTIAGILVALTFYSWALGGPAAADWHSRTAAVCVLTIAVVPYAAFLMWLALARA 2120
Qy 994 --VPNQGIINPM 1004
Db 2121 WMVP-QFTLVEPM 2132

RESULT 13
SLP1_CLOTH
ID SLP1_CLOTH STANDARD; PRT; 1664 AA.
AC Q06852;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Cell surface glycoprotein 1 precursor (Outer layer protein B) (S-layer
protein 1).
GN Name=olpB;
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE=93209311; PubMed=8458832;
RA Fujino T., Beguin P., Aubert J.-P.;
RT "Organization of a Clostridium thermocellum gene cluster encoding the
RT cellulosomal scaffolding protein cipa and a protein possibly involved
RT in attachment of the cellulosome to the cell surface.";
RL J. Bacteriol. 175:1891-1899(1993).
CC - SUBUNIT: Assembled into mono-layered crystalline arrays.
CC - SUBCELLULAR LOCATION: Cell wall.
CC - SIMILARITY: Contains 4 S-layer homology (SLH) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X67506; CAA47841.1; -.
CC PIR; T18262; T18262.
CC InterPro; IPR008965; Cellul_bind.
CC InterPro; IPR001119; SLH.
CC Pfam; PF003195; SLH; 3.
CC PROSITE; PS01072; SLH DOMAIN; 2.
CC Cell wall; Repeat; S-layer; Signal.

```

```

FT SIGNAL 1 28 Potential.
FT CHAIN 29 1664 Cell surface glycoprotein 1.
FT DOMAIN 36 763 4 X 156 AA approximate repeats.
FT REPEAT 36 191 1.
FT REPEAT 207 363 2.
FT REPEAT 409 565 3.
FT REPEAT 607 763 4.
FT DOMAIN 771 1377 approximate tandem repeats of T-P-S-D-E-
FT DOMAIN 1378 1449 P.
FT DOMAIN 1453 1494 Gly/Pro/Ser/Thr-rich.
FT DOMAIN 1495 1565 SLH 1 (incomplete).
FT DOMAIN 1566 1625 SLH 2.
FT DOMAIN 1626 1646 SLH 3.
FT DOMAIN 1664 1664 SLH 4 (incomplete).
SQ SEQUENCE 1664 AA; 178194 MW; 5F396695BA9PE74B CRC64;

Query Match 20.6%; Score 1108.6; DB 1; Length 1664;
Best Local Similarity 21.0%; Pred. No. 2.9e-14;
Matches 359; Conservative 118; Mismatches 371; Indels 864; Gaps 69;

Qy 4 KTLPIVLLLLLVFV-----IQOVSSQE 26
Db 6 KVLISILLTLLIISTTSVNMSPAEATPSIEMVLDTKEVHVGDVITATIKVNNIRKLAGYQ 65
Qy 27 LSCK-----GRCFSEFERGECDDAOCKKYDKCCPDYSEFCAEVKDKNKR 73
Db 66 LNIKFDPEVLQVPDPATGEEF-----DKSMP-----VNR 95
Qy 74 ----TKKXPTP-----KPPVUDE-----87
Db 96 VLLTNSKYGTPVAGNDIKSGIINFATGNNLTAYKSGIDEHTGIGIFKVLKQKNT 155
Qy 88 -----AGSGLDNGD-----FKVTTPTDTSTQHNKSVTSKPIITAKPINR 127
Db 156 SIRFEDTSLMGAISGTSLFDWDAETITGVEVIQPDLI-----VVEAEPL-----200
Qy 128 PSIPNSDTSKETSUTVNKETT-----VE 151
Db 201 -----KQASVALELDKTKVKVGDIIITATIKIENMKNFAGYQLNIKYDPTMLEAIE 250
Qy 152 TKETTITNKOT-----STDGKEKTS-----AKE 175
Db 251 LETGSAIAKRTWPVTGGTVLQSDNFGKTTAVANDVGAGIINFABAYSNTLKYRETGVAEE 310
Qy 176 TQSIKTSAKOLAPTSKVLAKPTPKAETTT-----KGPALTFP-----213
Db 311 TGIIGKIGFVLKAGSTAI-----RFEDTTMPGALEGTYMEDWYGENIKGYSVQVQGEI 365
Qy 214 -----KEPTPT-TPKEPASTTPKEPTPT-----235
Db 366 VAEGEPGEETPEEPVPTETPDVPTPTVTEEPVPSLPSYVIMELDKTKVKVGDIIITAT 425
Qy 236 -----TIKSAPT-----242
Db 426 IKIENMKNFAGYQLNIKYDPTMLEAIELETGSAIAKRTWPVTGGTVLQSDNFGKTTAVAN 485
Qy 243 -----242
Db 486 DVGAGIINFABAYSNTLKYRETGVAEEGIIIGKIGFVLKAGSTAIRFEDTTMPGAIEG 545
Qy 243 -----TPKEPATTTKSAPTTTPKEPATTTPKEPATTTPKEPATT-273
Db 546 TYMFDWYGENIKGYSVQVQGEIAGVEEPTPEPVT-----ETPVDPTPTVTEEPVSE 599
Qy 274 -----TPK-----276
Db 600 LPDSYVIMELDKTKVKEGDVIIATIRVNNIKNLAGYQIGIKYDKVLEAFNIETGDPIDE 659
Qy 277 -----276
Db 660 GTWPAVGVTILKNRDYLPVGAINNVSKGIILFAAYVYVDDYDEEGKSEDTGIIGNIGF 719
Qy 277 -----EPAP--TTTKEPATTTKS 293

```









GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 20.4912 Seconds  
(without alignments)  
4723.689 Million cell updates/sec

Title: SEQ1-E  
Perfect score: 5373  
Sequence: 1 MAWKTLPIYLLLLSVFVIQ.....DMDYLRVNPQGIINPMLS 1006

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79; \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1270.2	23.6	3020	2	A43932
2	1108.6	20.6	1664	2	T18262
3	1002.8	18.7	1274	2	T16251
4	990	18.4	1489	2	T31108
5	973.9	18.1	7962	2	I38346
6	969.9	18.1	2187	2	T30826
7	952.1	17.7	3570	2	T45025
8	892.3	16.6	1367	1	S48478
9	874.6	16.3	1188	2	S49915
10	845.4	15.7	6642	2	T29757
11	844.6	15.7	3507	2	T34513
12	844.3	15.7	1229	2	T25697
13	822.8	15.3	1151	2	T18535
14	822.5	15.3	1344	1	A35175
15	811.2	15.1	2897	2	B48666
16	811.2	15.1	3256	2	A48666
17	807.6	15.0	5762	2	A41819
18	757.4	14.1	1832	2	T31113
19	754.1	14.0	4135	2	T42629
20	750	14.0	2232	2	T34434
21	742.3	13.8	2142	2	B35098
22	736.8	13.7	3942	2	T42730
23	736.3	13.7	5262	2	T03454
24	735.6	13.7	761	2	C94672
25	735.4	13.7	990	2	I51618
26	733.5	13.7	924	2	S27923
27	729	13.6	971	2	T19431
28	728.5	13.6	4548	1	S00657
29	728	13.5	2225	2	T26063

## ALIGNMENTS

## RESULT 1

A43932  
mucin 2 precursor, intestinal - human (fragments)  
N;Alternate names: mucin SMUC-41  
C;Species: Homo sapiens (man)  
C;Date: 10-Mar-1993 #Sequence revision 12-Apr-1996 #text change 09-Jul-2004  
C;Accession: A49963; A45106; A43932; B33532; A61257; P00328; P00329  
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A;Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of the  
A;Reference number: A49963; MUID:94132002; PMID:8300571  
A;Accession: A49963  
A;Molecule type: mRNA  
A;Residues: 1-639 <GU1>  
A;Cross-references: UNIPROT:Q02817; UNIPROT:Q14884; GB:I21998  
R;Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A;Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both upstr  
A;Reference number: A45106; MUID:93016075; PMID:1400449  
A;Accession: A45106  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 626-1895 <GU2>  
A;Cross-references: GB:M94131; NID:G186395; PIDN:AAA59163.1; PID:G186396  
A;Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A;Accession: B45106  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 2037-3020 <GU3>  
A;Cross-references: GB:M94132; NID:G186397; PIDN:AAA59164.1; PID:G186398  
A;Experimental source: colon  
A;Note: sequence extracted from NCBI backbone (NCBIP:116698)  
R;Toribara, N.W.; Gum Jr., J.R.; Culhane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen, G.M.  
J. Clin. Invest. 88, 1005-1013, 1991  
A;Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polymorp  
A;Reference number: A43932; MUID:91358717; PMID:1885763  
A;Accession: A43932  
A;Molecule type: DNA  
A;Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A;Cross-references: GB:M74027; NID:G188863; PIDN:AAA59875.1; PID:G188864  
A;Note: sequence inconsistent with the nucleotide translation  
A;Note: sequence extracted from NCBI backbone (NCBIP:55749; NCBIP:55750)  
R;Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A;Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and eviden  
A;Reference number: A33532; MUID:89197956; PMID:2703501  
A;Accession: B33532  
A;Molecule type: mRNA  
A;Residues: 1916-2193 <GU4>  
A;Cross-references: GB:M22405; NID:G188873; PIDN:AAA36334.1; PID:G188874  
A;Experimental source: intestine  
R;Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.

30	721.6	13.4	801	2	T29018
31	719.8	13.4	1872	2	S36152
32	718.1	13.4	3938	2	T42761
33	716.3	13.3	1870	2	S37671
34	713.8	13.3	1630	2	A53577
35	712.1	13.3	1777	2	T34369
36	707	13.2	2774	2	A43359
37	698.4	13.0	1611	2	T38236
38	696	13.0	839	2	F75518
39	692.6	12.9	4957	2	T03455
40	687.4	12.8	1118	2	A48292
41	680.9	12.7	3381	2	T42389
42	679.3	12.6	856	2	T16543
43	674.9	12.6	5170	2	T15348
44	674	12.5	4006	2	T09070
45	671.5	12.5	3149	1	Q0BE8

hypothetical prote  
MHC class III hist  
Bassoon protein -  
MHC class III hist  
ascites sialoglyco  
hypothetical prote  
microtubule-associ  
hypothetical prote  
ALR protein - huma  
mucin, tracheobron  
versican precursor  
hypothetical prote  
hypothetical prote  
probable tenascin  
BPLF1 protein - hu

J. Clin. Invest. 87, 77-82, 1991

A;title: Human bronchus and intestine express the same mucin gene.  
A;reference number: A61257; MUID:91086401; PMID:1985113  
A;Accession: A61257  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 'V', 1925-1948, 'TTS', 1952-1954 <JAN>  
A;Experimental source: bronchus  
R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstner,  
Biochem. Biophys. Res. Commun. 183, 821-828, 1992  
A;title: Human intestinal mucin-like protein (MLP) is homologous with rat MLP in the C-C-  
A;reference number: PQ0328; MUID:92198477; PMID:1550588  
A;Accession: PQ0328  
A;Molecule type: mRNA  
A;Residues: 2328-2468 <XUG>  
A;Cross-references: GB:M86523  
A;Experimental source: small intestine  
A;Accession: PQ0329  
A;Molecule type: protein  
A;Residues: 2328-2342, 'K', 2344-2354 <XUG1>  
C:Genetics:  
A;Gene: GDB:MUC2  
A;Cross-references: GDB:I20203; OMIM:158370  
A;Map position: 11p15.5-11p15.5  
A;Superfamily: von Willebrand factor; von Willebrand factor type A repeat homology; von  
C;Keywords: glycoprotein; intestine; tandem repeat  
F:2766-2834/Domain: von Willebrand factor type C repeat homology <WVC>

Query Match 23.6%; Score 1270.2; DB 2; Length 3020;  
Best Local Similarity 21.1%; Pred.No.2.6e-18;  
Matches 390; Conservative 87; Mismatches 470; Indels 898; Gaps 56;

	3 WK-----	TLPIYILLLSVFVIQVSSQLSCKGRCFESF-----E 38	
Qy	:	: : :	
Dd	649 WRHVNVDVSCPNQSVFLNLTTCQTCSRSLSEADSHCLGEPAPDVGCGPDHTLDE 708		
Qy	39 RGR-----ECDC-----DAQKKYDKC-----		55
Dd	709 KGRCVPLAKSCYHRGLYLEAGDVVVROEBRCVCRDLRHLCRIQLIQGSTAPKIHMDC 768		
Qy	56 -----CPD-----YESFC-----		63
Dd	769 SNLTALATSKPRALSQTLAAGYYHTECVSGVCPCDGLMDDRGGCVVEKECPCVHNNDL 828		
Qy	64 ----AEVK-----DNK-----		70
Dd	829 YSSGAKIKVDNCCTCKRGWRVCTOAVCHGTCSIIYGSHYITFDGKYIDFGHCYSVAVQ 888		
Qy	71 -----KNRTKKPTPKPV-----		84
Dd	889 DYCGQNSSLGSFIITENVPCTGTGTCASKAIKFMGRTELKEDKRVVIQRDEGHVA 948		
Qy	85 -----VDEAGSL-----DNGDFKV-----		99
Dd	949 YTTREVGQYLNVVESTGIIVWDKRTTVFKIAPSYKGVTCGLCGNFDRHSNNDTTRDH 1008		
Qy	100 -----TPDTSSTQ-----ENKYST 114		
Dd	1009 MVVSSELDFGNSWEKAFTCPDYVTNPPEPCSLNPHRSWAEEKQSILKSSVFSICHKVD 1068		
Qy	115 SP-----		116
Dd	1069 KPFEACVHDSCSDTGDCFCPSASVAYAOECTKEGACVFWRTPLDLPICFYDYNPPH 1128		
Qy	117 -----KITTAKPIN-----		125
Dd	1129 ECEWHYPCGNRSFETCRTINGIHNSISVLYEGCYPRCPKDPRIYEELKKCVADKCG 1188		
Qy	126 -----PPSLP-----PNSD 135		
Dd	1189 CYVEDTHYPGASVPTTECTCKSCVCTNSQVVCPEEGKILNQTDGAFCYWEICGNGT 1248		
Qy	136 TSKE---TSLTVNKET-----TVETKETTTNKQSTDGKKTTSK-----		174

Db 2282 VEIVKVECEPPMPTCSNGLQVRVEDPDGCCWHECDYCTGMGDPRHYVTFDGLYYSYQ 2341  
Qy 968 GETPHMLLRPHVRPEVTPDMD----- 989  
Db 2342 GNCYVLVE-----EISPSVDNFGVYIDNVHCDPNDKVCSPRLIVRHETQEVLIKTIVH 2395  
Qy 990 -----YLRVPVNOGIIIN 1002  
Db 2396 MPMQOVQVNRQAVLPYKKGYLEVYQSGINYYVDIPELGVLVS 2440  
RESULT 2  
T18262  
S-layer protein - Clostridium thermocellum  
C:Species: Clostridium thermocellum  
C:date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18262  
R:Fujino, T.; Beguin, P.; Aubert, J.P.  
J. Bacteriol. 175, 1891-1899, 1993  
A:title: Organization of a Clostridium thermocellum gene cluster encoding the cellulose  
e.  
A:Reference number: Z18847; MUID:93209931; PMID:8458832  
A:Accession: T18262  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1664 <FUJ>  
A:Cross-references: UNIPROT:Q06852; EMBL:X67506; NID:g296879; PID:g296881; PIDN:CAA47841  
Query Match 20.6%; Score 1108.6; DB 2; Length 1664;  
Best Local Similarity 21.0%; Pred. No. 2.1e-15;  
Matches 359; Conservative 118; Mismatches 371; Indels 864; Gaps 69;  
Qy 4 KYLPYVLLLLLVFV-----IQQVSSOE 26  
Db 6 KVLISILLTLLIISTSVNMSFAEATPSIEMVLDTVEHVGDVITATIKVNNIRKLAGIQ 65  
Qy 27 LSCK-----GRCFSEFREGRECDCAQCKYDKCPCDYVESFCAEVKDNKNR 73  
Db 66 LNIKFDPEVLQVPDPATGSEFT-----DKSMP-----VNR 95  
Qy 74 -----TKKPTP-----KPPVUDE----- 87  
Db 96 VLLTNSKYGTPVAGNDIKSGIINFATGNNNTAYKSSGIDHTGIGIFKVLKKQNT 155  
Qy 88 -----AGSGLDNGD-----FKVTPDTS\*TOHKNVSTSPKITTAKPINR 127  
Db 156 SIRFEDTLSMPCAGISGTSLFDWDAETITGYEVIQDLI-----VVEAEPL----- 200  
Qy 128 PSLPNSDTSKETSITVANKET-----VE 151  
Db 201 -----KQASVALELDKTKVKVGDIIITATIKIENMKNPAGYQLNIKYDPTMLEAIE 250  
Qy 152 TKETTTINKQT-----STDGKEKTS-----AKE 175  
Db 251 LETGSAIAKRIWPVGGVTLQSDNYGKTTAVANDVGAGIINFABAYSNLTKYRETVGVAEE 310  
Qy 176 TQSEKTSAKOLAPSKVLAKTPKAETTT-----KGPALTPP--- 213  
Db 311 TGIICKIGFRVLKAGSTAI-----RFEDTAMPAGIEGTMYFMDWYGENIKGYSVVQPGEI 365  
Qy 214 -----KBPPTT-TPKEPASTTPKEPTPT----- 235  
Db 366 VARGEERPEERTEEVPTPTVDPTPTVTEBPVSELPDSYVIMELDKTKVKVGDIIITAT 425  
Qy 236 -----TIKSAPT----- 242  
Db 426 IKIENMKNFAGYQLNIKYDPTMLEAIELETGSAIAKRIWPVGGVTLQSDNYGKTTAVAN 485  
Qy 243 ----- 242  
Db 486 DVGAGIINFABAYSNLTKYRETVGVAEETGIICKIGFRVLKAGSTAIRPEDITAMPGAIEG 545  
Qy 243 -----TPKEPAPTTKSAPTTPKEPAPTTKKEPAPT- 273

Db 546 TYMPDWYGENIKGYSVVQPGEIIVABGEETPEPVET-----ETEVDPETPTVTEEPVSE 599  
Qy 274 -----TPK----- 276  
Db 600 LPDSYVIMELDKTKVKEGDVIIATIRVNNIKNLAGYQIGIKYDKVLEAFENIETGDPIDE 659  
Qy 277 ----- 276  
Db 660 GTWPAVGGTILKNRDYLTGTGAINNVSKGILNFAAYVYFDYREGEKSEDTGIIGNIGF 719  
Qy 277 -----EPAP-----TTTKBPAPTTTKS 293  
Db 720 RVLKAEDTIRFELESMPGSDIGTWLMDWLNIRISGVVVIQAPAKADEIPTDTTPS 779  
Qy 294 -AFTTPKEPAPTTTKKPAPTTTPKEPAPTTTPKEPTPT-TPKEPAPT-TKEPAPTTTPKEPAP 350  
Db 780 DEPTPSDEPTSD-----EPTPSDEPTPS-DEPTSETPEEPIPTDTPSDEPTPSDEPTP 833  
Qy 351 TAPKKA--PTTPKEPAPT-TPKEPAPTTTKPSPTTPKEPAPTTTKSAPTTPKEPAPTT 407  
Db 834 SDEPTPSDEPTPSDEPTSETPEEPIPTDTPSDEPTPSDEPTPS-DEPTPSDEPTSD 890  
Qy 408 TKSAPTTPKEPSPTTPKEPAPTTTPKEPAPTTPKKA--PTTPKEPAPT-TPKEPAPTTTK 464  
Db 891 EPTPSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTSETPEEPIPTDTP 950  
Qy 465 KPAPTAPKEPAPT-----TPKETAPTTPKKTPT--TPEKLAPTTPKEPAPT-TP 511  
Db 951 SDEPTPSDEPTPSDEPTSD-ETPSDEPTPSDEPTSE-----TPPEPIPTDTP 1005  
Qy 512 BELAPTTPPEPT-----PTTPKEPAPT--TPKAAAPNTTPKEPAPT-TPKEPAPTTPKKA 563  
Db 1006 SD-EPTPSDEPTPSDEPTPSDEPTPSDEPTP-SDEPTPSDEPTPSDEPTPEEPIPTDTPSDE 1063  
Qy 564 PTPPKETAPT--TPKGTAPTTLKBPAPT-TPKKAPKEL--APTTPKEPAPTTSDKPA 616  
Db 1064 PTPSDEPTPSDEPTP-SDEPTPSDEPTSETPEEPIPTDTPSDEPTPSDEPT--PSDEPT 1120  
Qy 617 PTPPKGTAPTTPKEPAPT-TPKEPAPTTTPKGTAPTTLKBPAPTTPKKA----- 664  
Db 1121 PS-----DEPTPSDEPTSETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTP 1176  
Qy 665 ---PKELAPTTTKGTSTTSOKPAPT--TPKETAPTTPKEPAPT-TPKKAPTTTPETPP 717  
Db 1177 SETPEEPIPTDTPSDEPTPSDEPTPSDEPTPSD-EPTPSDEPTPSDEPTPEEPIPTDTPSDE 1235  
Qy 718 PTTSEVSTPTTKETPTTIHKSPDESTBELSAEPTPKALENSPKERGVPVTKTPAAKPEM 777  
Db 1236 PTPSD--EPTPSDEPT--PSDEPTP--SDEPTP--SETPEEP-IPTDTPSDEPTPSD 1283  
Qy 778 TTTAKDKTTERDLRTPPTTTAAPKMTKETATTTTEKTTESKITATTTQVSTTTTQDTTFF 837  
Db 1284 EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTSETPEEPIPTDTPSDEPTPSDEPTPS 1343  
Qy 838 KITLKITTLAPKVTITTKITTTTEIMNKBEETAKPKDRATNSKATTPKQ-KPTKAPK 896  
Db 1344 DEPT-----PSDEPTPSDEPTPSDEPTPSDEPTSETPEE 1378  
Qy 897 PTTSTKPKKTPMRVKKPTTPTPKMTS-----TMPELNPSTR 933  
Db 1379 PT-----PTTPTPTSTPTSGSGSGSGSGGGGGTPTVPTSPPTPTS- 1424  
Qy 934 IAEAMLOTTTRPNQTPNSKLVEVNPKSEDAGAEGETPHMLLRPH--VFMPE----- 983  
Db 1425 -----KPTSTAPTEIE-EPTPSDVPGALGGEHRAVLRGYPGSGFAPENITRAE 1473  
Qy 984 -----VTPD----- 987  
Db 1474 AAVIFAKLLGADESYGASPSYDLADTHAAWAIAKFATSOGLFKGYGPDGTFKPDQNT 1533  
Qy 988 -----MDYLRPVNPGI-----IINP 1003

Db 1534 RAEFATVVLHFLTKVKGQIMSKLATIDISNP 1565

RESULT 3  
Tl6251  
hypothetical protein F35A5.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: Tl6251  
R:Leimbach, D.  
submitted to the EMBL Data Library, January 1996  
A:Description: The sequence of C. elegans cosmid F35A5.  
A:Reference number: Z18485  
A:Accession: Tl6251  
A:Molecule type: DNA  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-1274 <LET>  
A:Cross-references: UNIPROT:Q20007; EMBL:U46675; NID:gl166613; PID:gl166621; PIDN:AAB526  
A:Experimental source: strain Bristol N2; clone F35A5  
C:Genetics:  
A:Gene: CESP:F35A5.1  
A:Map position: X  
A:Introns: 1272/2

Query Match 18.7%; Score 1002.8; DB 2; Length 1274;  
Best Local Similarity 23.1%; Pred. No. 2e-13;  
Matches 330; Conservative 99; Mismatches 309; Indels 692; Gaps 75;

QY 74 TKKKPTP-----KPP--VDEAGSLDNGDFKVTTPDTSTQHKNVSTSPKITTAKP 123  
DB 2 SRAPPTPIKNPAKKKPPWESVDE----- 25

QY 124 INRPRLPNNSDTSKETSIVNKTETVETKTTTINKQTSIDGKKTSAKETQSIKTS 183

DB 26 -----EEMEVEDEETAPSK-----LEKKPSLKR----- 49

QY 184 AKDLATSKVLAKPTPKA-----EYTTKGPAITTP-----KEPTP- 218  
DB 50 -KD-APTQPV--FSGAPSFVPIKNPVKKWKAPWEDDEMEEAAPAPVAKKVRDPSK 104

QY 219 TTPKEPASTTPKE-----PTPT-----TIKSAPTTP 244

DB 105 KVPKPRDAGSPKIMAAKKEPELIPVPTFVKNPVKKFKAPWEDDEVDVVDKDAFTVP 164

QY 245 -----KEPA-----PTTKSAPTTPKSP--APTTPKEPA-----PTT 274

DB 165 AKKTPVLKKKEPAAAKPRDPSPKAAPSKEHDPIVPTPIKNPAKKWKPPWEDDEVPT 224

QY 275 P-KEPAPTTPKEPA-----PTTKSAPT----- 296

DB 225 EIKSEPEATRKVPALKKKKSPSTSVKVSDESPSTKKVPVKKKEPEVPTPIKNPTKKKKPPW 284

QY 297 --TP-----KEPAP-----TTPKKPAPTTPKEPAPTTPK 323

DB 285 EDETPVEVKEPPVPEKAPVLKKDPAKARDPSPKAAKPKVESPSPVVP--PTPVK 343

QY 324 EPT-----PTTPKEPAPTTPK-----EPAPTTP-KEPAP----- 350

DB 344 NFVKYKPPWVDDDEPAEEVKKPSAEPKTPVLKRKEPEPSSTTPSSDPSPKAAAPAVK 403

QY 351 --TAPKKPAP-----PTTKKEP-----TTPKEPAPTTP-- 369

DB 404 RDSGPKATPLQADPKAQEVPTFVKNPVKKYPWEDDEDPVEEVKQEPAPAKKTPVL 463

QY 370 --KEPA-----PTTKKEP-----GPTTP-KEPA----- 389

DB 464 KRKEPAKOTAKPATSKTPTPEIKKDPVKPRDSGPKVAAKPDGAQAPATPVKNPVKKWR 523

QY 390 -----PTTKKSAPT--TTKEPAP-----TTTKSAPTTPKEPSPTTKKE 425

DB 524 PPWEDDETADDVSKPTDAKKTPTSLAKKDPAPAKESLKPADTK-APAKPRDPSG---KK 579

QY 426 PAPTTPKEPAPTTPPK-----PAPTTPKEPAPTTPKEP-----APTTPKK 465

Db 580 VAPTAPEKKTVPVLAKKEPAGPADSKTKEPESKPRDPSFKKAVPAKVPKTEVAPAAVKK 639

QY 466 PAP-----TAPKEPAPTTPKETAPTTP-----KKLTP-----TTPKELAP 500

DB 640 PEPISKPDXTAPKKAEPNSP--VVPPTPVKNPVKKWKPPWEDDDAPAKPVSLPEEKKTP 697

QY 501 TTPKEPAPTTP-----ELAPTTPPEEPTP--TTPPEEAPTTP----- 534

DB 638 VLAKK-APTKEPDESEAAADPVSGSPSSKDPKLAKKAFVKPRDPSPMKAVPIKPAKTEVPPA 756

QY 535 -----PK--AAAPNTPKBPAPTTPKEP----- 554

DB 757 VVKKEPVPVAKSRDPSPKKAEPNSPVVP--PTPVKNPVKKWKPPWEDDDAPAEVNVVPEP 815

QY 555 -----APTTPKEPAPTTPKETAPTTPGTATPTTLKEPAPTTP--PKKPAPKE----- 598

DB 816 EKKTPVLAKKTPVKPRDPSFKKAVPAKPSSTKTDDAPVSVKKPEPVSKPEPSPKKAEPNS 875

QY 599 --LAPTTPKEPT-----STTSKPAAPTTPKTAPT--TPKEP-----APTTPKEP 639

DB 876 PVVPTFVKNPVKKWKPPWEDDEPTEEVKPSSEPKKTPVLAKKEPEKPKDAPKAAKP 935

QY 640 APTTPKGTAPTTLKBPAPT-----TPKKAP-----KELAPTTPKGTTS-----TTS 681

DB 936 RDPSPKKAEP--KEPAKVAAPKPRDLSPKKAIPIPANTQEAAPTTPVKNPVKKWKPPWEDD 993

QY 682 DKP-----APTTPKEPAP-----TTPKKPAPTTPPETPTTPPTTSEVSTP 726

DB 994 DEPAEVSAPKEPKTPVLAKKAPKAPRDPSPKKAAPVAAKDPKPIPEVP-----P 1044

QY 727 TTTKEPTTIHKSP-----DESTPELSA-BP--TPKALENSPKPEGV-----P 765

DB 1045 TPVKNPVKKWKPPWEDDDDESPVSAPEPEKKTTPVLAKKAPTTPATKPDSEAAADPVSGP 1104

QY 766 TTKTPAAT-----KPEMTTTAKDKTTERDLRTTPETT--TAAPKMTKETATTTTEKTES 817

DB 1105 TSKDPKLSKKAPEKPKPTTDPKDDKLKPSPAKKKEPAPEAPAAKPKWKPPWDDDDPDEPEA 1164

QY 818 KITATTQVTSITTTQDTPPKITTLTTLTTLAPKVTTTKITTTTTEIMNKPEETAAPKDDA 877

DB 1165 DFTVPAPSKKPDTEADPLG-----GPKTKDPK-----LNKKAPEKPTPK- 1206

QY 878 TNSKATTPKPKPTKAPKPTSTKPKTMPRVKRPK-----TTPTPRKMSTMT 925

DB 1207 -----PKPEKVSKEPPKTEPPKP--AAPKKWKPPWEDDDEPEADFTMPAPKK----- 1253

QY 926 PELNPTSRIAEAMLQTTTRPNQTPNSKLVEVNPKESED-----AGGAGEPTP 971

DB 1254 -----PDTEPDADSLGGPKPKDP 1271

## RESULT 4

T31108  
cyst germination specific acidic repeat protein precursor - Phytophthora infestans  
C:Species: Phytophthora infestans (potato late blight agent)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T31108  
R:Goernhardt, B.  
submitted to the EMBL Data Library, April 1998  
A:Reference number: Z20986  
A:Accession: T31108  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1489 <GOE>  
A:Cross-references: UNIPROT:O96449; EMBL:AF061185; NID:g3851513; PID:g3851514; PIDN:AAC8  
C:Genetics:  
A:Gene: car90

Query Match 18.4%; Score 990; DB 2; Length 1489;  
Best Local Similarity 26.1%; Pred. No. 4.3e-13;  
Matches 391; Conservative 53; Mismatches 415; Indels 640; Gaps 70;







```

QY 795 ETTAAAPKMTKATTTTEKTTESKITATTQVSTTTTODTTPTEKITLITLITLAPKVTIT 854
Db 7294 VKVPEAPK-----EWPEKKV-----FVPPPKKEVPTTKVEV-----PKVAVP 7333
QY 855 KKTITTT-----EIMNKPETA----- 871
Db 7334 EKKVPEAIPPKPEPPPEVPEEPEVALEPPAEVVERPEPAAPPQVTVPPKNFVPEKKA 7393
QY 872 ----- 903
Db 7394 PAVVAKKPELPPVKVPEVPEVPEKKVPLVVPKKPEAPPAKVPEVPEVPEKKVAVPK 7453
QY 887 -----POKPTKAPKKPTSTTKP----- 903
Db 7454 KPEVPEAKVPEPKPVLEEKPAVVPVERAESPPEVPEEPEETAPBEEETAPBEKPVV 7513
QY 904 -----KTMPEVRKPKTTP----- 916
Db 7514 AEEEEVPPPAVPEPKKIIPKPKVPIKPEAPPPEPEPEKPKLKPBPPEPP 7573
QY 917 -----TPRKMTS----- 923
Db 7574 APBKEDVKEKIFOLKALPKKVPENPOVPEKVELTPLKVPGGKKVKLLPERKPEKEE 7633
QY 924 -----TMPELNPTSRIAEAMLOTTTR-----PNQ 947
Db 7634 VVLKSVLRKRPEEPEKVPKLEKVKKPAVPPPPPKPEVEEVPTVTKREKIDPEPK 7693
QY 948 TPNSK-----LVEVNPKSDAGAGETPHMLLRHVFVPEVTP----- 986
Db 7694 VPEIKPAIPLPAPEPKPE-----AEVKT-----IKPPVPEPPTIAAPVTVPVVGKAE 7745
QY 987 -----DMDYLPVRVNOGI 999
Db 7746 AKAPKEAAKPGPIKGV 7763

RESULT 6
T30826
N:ascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30826
R:Yotcov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A>Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle
A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:g1666688; PID:g1666689; PIDN:AAH187
A:Gene: Naca
A:Map position: 10
A:Iron: 24/1; 1996/1; 2024/3; 2050/3; 2099/3; 2142/3; 2183/3
A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 18.1%; Score 969.9; DB 2; Length 2187;
Best Local Similarity 21.7%; Pred. No. 1.8e-12;
Matches 351; Conservative 134; Mismatches 412; Indels 721; Gaps 67;

QY 4 KTLPTLYLL-----LISVFIQOVVSQELSKGRCFESPERG 40
Db 479 KNLPIISALVNGAVPSVPAQAGLPTKDTTLOPLAIKESFSSQAS-----SLEVL 531
QY 41 RECDCAOCK-----YDKC----- 55
Db 532 SE---DVTVKITGGPVPVVPATAGVATTTSLRADSPAVIRADSCVSPNTVQPLKRS 588

```

```

QY 56 -----CPDYBSFCAEVKDNKKNR-----KKX 77
Db 589 VTDPAAPRTAKTAPSTTSPVLPLASBGCVPASSMALSPOVASVSETALALSPPIKSV 648
QY 78 PTPKPPV-----VDEAGSLDNGDFKVTTPTDSTTOH----- 109
Db 649 PFPDPPLAEISFSNARKVDVAHVHMESSGSRQGHDPASVTAKGTWVCLADSLDTSVSAS 708
QY 110 -----NKVSTSPKITTAKPINPRPSLPNSD 135
Db 709 KGSALSGASSPLYLEVFLPEAGLAVQGPKGSLNKLSPTPSSKGAHV-----PSTGAPP----- 764
QY 136 TSKEISUTYAKETTVETKE-----KDLAPTSKVLAKP-----TTTTN 159
Db 765 -SPKGAPIVPTSESISSKQVPAEILPSPQKTPPEVTASRLISAVQSPKVDPIMSDVTITSP 823
QY 160 KOTSTDGKEKTTSAKETQSIETKSA-----KDLAPTSKVLAKP----- 197
Db 824 KKTISATAVPKOTSA--TLSLSKVPATVLSLPPKAPVAPSNEATIVPTETSLKNALAAA 881
QY 198 -----TPKAEITTKGPALTT-----PKE----- 215
Db 882 TPKEITLATSIPKVTSPSPQKTPKSVLSKGAAMTSKATEIAASKOVSPQFPKEVPLLQ 941
QY 216 --PTPTPKPEASTT-----PKEPTPTTIKSAPTTPKE--PAPTITTKSAPTTPK-- 260
Db 942 HVPEPTSPKSPVSDTLGALTSPPKGP--PATLAETPTVPKSPKPAASKKTATPESPEG 1000
QY 261 -----EPAPTITTKPEAPTTPKEAP-----T 289
Db 1001 VTAVPLEIIPCCKAKPAKTAAPKESATSSSKRAPKXTAVSKBPSKGVTAVPLEISLPKE 1060
QY 290 TTKSAPTTPKEAPTTPKK--PAPTTPKE-----PAPTTPKEPTP----- 327
Db 1061 TSKSA--TPGEKSASSPKSPKTAGPKETPPGGVTAVPPEISLPPKPTFONATPNSLAA 1118
QY 328 -----TTPKE-----PAPTTPKE-----APTTP 345
Db 1119 SSQKSGSPKTSVPKETPPGGVTAAPLEIPSAQKAPKTAVPKQIPTPEDAVTILAGSLSP 1178
QY 346 KEPAETAPKPKAPTTP-----KBPAPTTPKE 371
Db 1179 KAASKTAAPKEAPATPSPGVIAVSGEISPSPKKTSKTAAPKENSATLPKRSFKTAAPKE 1238
QY 372 PAPTTPKE-----PSPTTPKE--PAPTITTKSAPTTPKEAPTTPKSAPTTP--PK 416
Db 1239 -TPATSSSGVTAVPSEISPSPTTPASKGVVTLTPKGAPNALAE--SPASPKVKVKTAAPE 1296
QY 417 EPS--PTTTPKEAPTTPKEAPTTPKKAAPT-----PKEAPTTPKEAPTTPTKKP 466
Db 1297 ETSPTTPSQKIPKVAGPKESATPSPKKTPTKTAVPKETSAPSEGVTAVPLEIPPSPRKAP 1356
QY 467 APTAPKE--PAPTTPKETAPTTPKLTPTTPKLAAPTTPKEAPTTPPELAPTTPPEPTPT 525
Db 1357 KTAAPKETPAPS--PEGATTAPVQI--PPSPRKGSKAGSKETPTTP-----SPEGVTA 1407
QY 526 TPEBP--APTTPKAAAP-----NTPKEAPTTPKEAPTTPKEAPTTPKEAP 564
Db 1408 PLEIPISSKTSKMASPKETLVTPSSKLSQTVGPKETISLEGATAVPLEIPPSHKKAPKT 1467
QY 565 TTPKETAPTTPKGAPTTPKKEAPTTPPKKAPKELAP-----TTTKEPTSTTSKPAPTTP 620
Db 1468 VDPKQVPLTPSKDAPTTLAE--SPSPKPK--APKTAAPPSESVTTVPP-----EKPA--TPQ 1519
QY 621 KGTAPTTPKEAPTTPKEA-----PTTPKGTAPTTPKKEAPTTPKPAKPAKELAPT 671
Db 1520 KASGTTASKVPVPAETQEVAVSSRETPTVPAVPPVKXPSSHKTSKTELKEADATLPPS 1579
QY 672 TTGPTSTTSKPAAPTTPKETAPTTPKKEAPTTPKKEAPTTPPETPTTPETVSPPTTKE 731
Db 1580 PTKSPKIPSSKK--APRTSAP-----KEFPASPSIKPVTTLSLAQT-----A 1618
QY 732 PTHHKSPEDESTPE--LSAETTPKALENSPKBFGVPTTKTPAATKPEMTTTAKDKTTERD 789

```

Db 1619 PPSLOKABSTIPKENLAA---PAVLPSVSSKSPAAP-APASASLSP---ATAAPOTAPKE 1671  
QY 790 LRTTP-----ETTTA-----APKMTKEATTTEKSKITATITQVSTTTQ 832  
Db 1672 ATTIPSCKKAATEPIETSTAPSLEGAPKETSEIV-----SKVLSPPKKAASSK 1724  
QY 833 DTTPPKITTL-----KTTTLAPKVVTTTKKIIITTEIMNKPEETAKPKDRATNSKATTPKPQ 888  
Db 1725 RASTLPALTPSLKASVLSPTSSGK-----DSHISPVSDACSTGTTT--PQ 1771  
QY 889 KPTKAPK-----PSTNKKPKT 905  
Db 1772 ASEKLPKKGPTAFTEMLAAPAPESALAITAIPQSPGANSNSASPKCPDPSSKDKYK 1831  
QY 906 MPR--VRKEKTTPTPRKMTSTWPELNPTSRISAEAMLQITTRPNQ----- 947  
Db 1832 LPSAVALAPQTPVEK-----DTSKAIETLLVSPAKGSDCLHSPKGPVGSQVATP 1881  
QY 948 -----TPNSKLVEVNPKSEDAAGAGETPHMLL 975  
Db 1882 LAAFTSDKVPPEAVSASVAPKAPAPASLTLAPSPVAPLPKQPLLESAPGVLESFSL- 1940  
QY 976 RHVFVMEVTPDMYLPV-----PNQGIILN-----PML 1005  
Db 1941 -----PVPABEDELPLIPPEAVSGGEFFQPILVNMPAPKDPAGTAPAPSAKQPV 1991  
RESULT 7  
T45025  
mucin MUC5B, tracheobronchial [imported] - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T45025  
R;Desseyn, J.L.; Guyonnet-Duperat, V.; Porchet, N.; Aubert, J.P.; Laine, A.  
J. Biol. Chem. 272, 3168-3178, 1997  
A;Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat  
A;Reference number: Z22899; MUID: 97166151; PMID: 9013550  
A;Accession: T45025  
A;Status: preliminary; translated from GH/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-3570 <DES>  
A;Cross-references: EMBL: Z72496; MID: g1834502; PIDN: CAA96577.1; PID: g1834503  
A;Experimental source: placenta  
C;Genetics:  
A;Gene: MUC5B  
Query Match 17.7%; Score 952.1; DB 2; Length 3570;  
Best Local Similarity 13.0%; Pred. No. 7.8e-12;  
Matches 426; Conservative 116; Mismatches 412; Indels 2319; Gaps 96;  
QY 6 LPYLLLLLVFVYQVSSQBLSCK-----GRCFESFER----- 39  
Db 3 LPV-----STVCREV-----CRMSSWYNGHRPEPLGGDDPETENLRQRCYQVCPV 50  
QY 40 -----GRECD-----AOCKYDKCCDDYE--SECAEVKONKK 71  
Db 51 LADIECRAAQLPDMPLBELGQVQVDCDRMGLMCANSQQSPPLCHDYELVLCCEY----- 105  
QY 72 NRTKKKTPKP-----PVDVAGSGLDN----- 82  
Db 106 --VPCGSPAPGTSQPQSLASAPVPTPTQTATEKTLTWTPSIRSTAALTSQTGSS 163  
QY 83 -----PVDVAGSGLDN----- 94  
Db 164 SGPVTVPSAFTTTCQRCQWTEFDEYDKSEQLGGDVESYDKIRAAGHLCOQPKDI 223  
QY 95 -----GDFKV----- 99  
Db 224 BCQAESFPNWTIAQVQKQVCHDVHGLVCRNWEQGVFKMCMYKRVIRVLCSDHCRGA 283  
QY 100 TTP-----DSTTQHNKV-----STSPKITTAKPINP----- 126

Db 284 TTPPTTELEATTTTQALFSTPQFTSSPGLTRAPPASTTAVPTLSEGLTSRVTSTLG 343  
QY 127 -----RPSLPN----- 133  
Db 344 TATTGGPQSGSTEPTVPGVATSTLPTRSALPGTGLGTRWPSQPTPLAPTTWATSRA 403  
QY 134 -----SDTSKE--TSL--TVNKETTVETKTTTTNKQTSQSDGKEKTTSAKET----- 176  
Db 404 RPTGASTASKEPLTSLAPLTLSELSTQASTSPRTETTMSPLTNITTSQGTTRCQPK 463  
QY 177 -----QSIE----- 180  
Db 464 CEWTEWFOVDFTPSGVASGDMETFENIRAAAGKCMWAPKSIIECRANVEPVIDQGVQL 523  
QY 181 -----KTSAKDLAP-TSKYL 194  
Db 524 TCSLETGLTCRNEQDGRNMCNVRVLCDDYSHCEPSTLSTSTSPSTPGTTWIL 583  
QY 195 AKP-----TPKASTTKGPAITTPKEPTP----- 218  
Db 584 TKPTTTATTASTGSTATASSTQATGTHVSTTATTPTVTSSK-ATPSSFGTATALPA 642  
QY 219 -----TPKEPASTTPKEPTTIK----- 238  
Db 643 LRSTATTPTATFTAIPTSSSLGTTWRLSQTTTPMATMTATPSSPTPEVHSTVLTTA 702  
QY 239 -----SAPTTT----- 244  
Db 703 TTTGATGVATPSTPGTAHTTKVLTITTTGTATPSSSPGRARTLPVMTITTTPTTRG 762  
QY 245 -----KEPAPTTT----- 252  
Db 763 STVTPSSIPGTHPTVLTITTTTATGSMATPSSSTQTSPTGTPSLTITATTATGSTT 822  
QY 253 --KSAP--TPKPEPAPTTT-KEPAPT---TPKEPAPTTPKEPAPTTKS-----APTTP 298  
Db 823 NPSSSTPGTPIPPVLTATTAPATSTVTPSALGTHTPPVNTTATTHGSLSDSSP 882  
QY 299 -----KEPAPTPKKPAPT---PKPEPAPTTPKEPTTTPKEPA 334  
Db 883 HTVCTAWTSATSGILGTHITEPSTGTSHTPAATGTTQHSPTALSSPHSPSSKTTESPPS 942  
QY 335 PTTKEPAPT--TPKEPAPTAPKAPAPT--PKBP--APT----- 368  
Db 943 PGITPGHTTATSKTTATATATPSKTRTSTLLPSQTSAPITVVVTMGCEPOCAWSEWLDYS 1002  
QY 369 ----- 368  
Db 1003 YPMPESGGDFDTYSNIRAAAGAVCEQPLGLECRAQAGVPLRELQGVVECSLDGLVC 1062  
QY 369 -----PKPEPAPTTPKES----- 381  
Db 1063 RNREQVKFKMCFNYEIRVFCNCGHCPSTPATSSATPSTPGTWTILTELTTATTE 1122  
QY 382 -----PTTPKEPAP-----TTTKSAPT----- 398  
Db 1123 STGSTATPTSLRTAPPKVLTTTATTPTVTSSKATPSSSFGTATALPALRSTATTPTAT 1182  
QY 399 ----- 398  
Db 1183 SVTPISSSLGTTWRLSQTTTPTATMTATPSTPETAHTSTVLATATTGATGVAT 1242  
QY 399 -----TKPEPAPT-----KSAPTTKPEP-----SPTT----- 422  
Db 1243 PSSTPGTAHTKVTITTTGTTATPSSSPGATLPVPMISTTTTPTTRGTVTPSSIPGT 1302  
QY 423 -----TKPEAPTTKPEPA 435  
Db 1303 THTATVLTITTTTATGSMATPSSSTQTSPTGTPSLTATTATTATGTTNPSPTGTRPI 1362  
QY 436 P-----TPPKKAPT---TPKEPAPTTPKEPAPT----- 463  
Db 1363 PPVLTATTATTPAATSSVTTPSALGTHTPPVNTTATTHGSLSPSSPHTVRTAWTSAT 1422

```
QY 464 -----KXPAPT-----APKEPAPTP----- 479
Db 1423 SGTILGTHITPSNGTSHTPAATGTTQHSPTALSSHPSSRTTESPPSGTTPGHHTA 1482
QY 480 --KETAPTTPK-----LTPTTPKLAAPT----- 503
Db 1483 TSRTTATATPSKTRTSTLLPSSPTS-APITTVTMGCEPQCAWSEWLDYSYMPGPGSGD 1541
QY 504 -----EKP----- 506
Db 1542 FDTYSNIRAGGAVCEQPLGLECRAQOGVPLRELQGVVECSLDGLVCRNREQVGKFK 1601
QY 507 -----APTPEELAPTPEB-----PTP 524
Db 1602 MCFNVEIRVFCNCGHCPSPATSSSTATPSPGTTWILTEQITAAATTAATGTAIPSS 1661
QY 525 TTPEEPAP-----TTP-----KAAAPNTPK-----EPAPTT-- 550
Db 1662 TPGTAPPKVLTSQAATFTATSSKATSSSPRTATTLPVLTSTATKSTATSFPIPSSTL 1721
QY 551 -----PKEP-----APTTPKEPAPTPKETAPTT-- 574
Db 1722 GTTGTSONRPHPMATMSTIHPSTPETHSTVLTAKATTATATSSMSTPSPGTTWI 1781
QY 575 -----PKGAPT-----LKEPAPT----- 589
Db 1782 LTELTAATTAALPHGTSPSTPGTTWILTEPSTATVPTGSTATASSTRATAGTLKV 1841
QY 590 -----TPKK-----PAKELAPT----- 602
Db 1842 LTSTATTPVISSRATPSSPGTATALPALRSTATTPTATSVTAIPSSSLGATWRLSQT 1901
QY 603 -----TTKEBTSTT----- 611
Db 1902 TTPATMSTATPSTPETHSTVLTATTTRTGVATPSSPGTAHTTKVPTTTTGTG 1961
QY 612 -----SDKPAPT----- 619
Db 1962 TATPSSSPGALTAPPVWISTTTTPTTRGSIPTPSSIPGTHATVLTITTTTIVATGSMAT 2021
QY 620 -----PKGAPTTPKEPAPT----- 635
Db 2022 PSSSTQSGTPSPLATTATITATGSTINPSSPGTTPPPVLTATTTPAATSVITPS 2081
QY 636 -----PKEP-APT-----TPK 645
Db 2082 SALGTHTPPVNTTATTHGRSLPPSSPHVTPTAWTSAATSGILGTHITPSTGTSHTPA 2141
QY 646 GTAPT-----LKEFAP-----TTPKKPAKELAPTTTKGTSTTSKPAPTTPKET----- 692
Db 2142 ATTGTTQSTPALSSHPSSRTTESPPSP-----GITTPGHTGRSTTATATPSKTRTST 2197
QY 693 -----AP-----TTPKEP----- 700
Db 2198 LLPSSFTSAPITTVVTTGCEPQCAWSEWLDYSYMPGPGSGDFTYSNIRAGGAVCEQP 2257
QY 701 -----AP----- 702
Db 2258 LGLECRAQOGVPLRELQGVVECSLDGLVCRNREQVGKFKVCFNVEIRVFCNCGHCP 2317
QY 703 TTPKKPAPTTPPTPTT----- 720
Db 2318 STPATSSATPSSTPGTTWILKLITATTTESTGSTATPSTQGPAGTHVSTTATTP 2377
QY 721 -----SEVSTPTTT----- 729
Db 2378 TVTSSKATPSSPGTATALPALRSTATTPTATSFALPSSSLGTTWRLSQTITPMATNS 2437
QY 730 -----KEPTTIHKS----- 738
Db 2438 TATPSSPTPETHSTVLTATTGATGSAVATPSSPTGTAHTTKVPTTTTGTFTVTPSSS 2497
```

```
QY 739 -----FDEST-- 743
Db 2498 PGTARTPPWISTTTTTPTTSGSTVTPSSIPGTHHTPTVLTITTTQPVATGSMATPSSSTQT 2557
QY 744 -----PELSAEPPTPKALEN----- 757
Db 2558 SGTPPSSLITATTATTATGSTTNPSSTPGTTPPIPELTITATTTPAATSTVTPSSALGTH 2617
QY 758 -----SPKEP-----GVPT 766
Db 2618 TTPVNTTATTHGRSLSPSSPHVTVAWTATSOTGLGTHITPSTGTSHTPAATTTGTT 2677
QY 767 TKTPAAATKPEMT-----TTAKDKTTE----- 787
Db 2678 TSTPALSSHPSSRTTESPPSGTTPGHTTATSRITATATATPSKTRTSTLLPSQTSAPI 2737
QY 788 ----- 787
Db 2738 TTVVTTGCEPQCAWSEWLDYSYMPGPGSGDFTYSNIRAGGAVCEQPLGLECRATAQP 2797
QY 788 ----- 787
Db 2798 GVPLGELQGVVECSLDGLVCRNREQVGKFKMCFNVEIRVFCNCGHCPSTPATSTAMP 2857
QY 788 -----RDLRTTPET-----TAAPKMTVKETA-----TTTEKTIESKI-- 819
Db 2858 SSTPGTTWILTELTTTATTASTGSTATPSSSTPGTAPPKVLTSPTATTPTATSSKATSSS 2917
QY 820 ---TATTQV-TSTTTQDT---TPFKITLLKTTLAPKVTTTK-KTITTEIMNKPEE- 869
Db 2918 SPTATLPLVLTSTAKSTATSVTPSPSLGTLGTLPSQTTTPPVATMSTIHPSSPTT 2977
QY 870 -----TAKPKORATMSKAT-----TPKPQK----- 889
Db 2978 HTSTVLTATKATRATSTSTPSTPGTTWILTELTAATTAGTGTATPSSPTGTTWL 3037
QY 890 ----- 889
Db 3038 TELTTTATTASTGSTATSLSTPGTTWILTEPSTTATVAPPSTATASSTQATAGTPHV 3097
QY 890 -----PTKAPKPTSTKKPKT---MPVRVKPKTTPTPRKMTSMPELPTS-----R 933
Db 3098 SITATPTVTSSKATPSSPGTATALPALRSTATTPTATSFAT--PSSSLGTTWTR 3152
QY 934 IABAMLQTTTRNPQTNPSKLVEVNPKSADAG-----GAEGTPEHMLLRPHVEMPE 983
Db 3153 LSQTTTPTATMTGSTATPSS---TPETVHTSTVLTITATTGATGS-----VATPS 3198
QY 984 VTPDMDYLPRVP-----NOGIINP 1003
Db 3199 STPGTAHTTKVPTTTTGTATPSSSPGALTLP 3231
```

## RESULT 8

S48478

glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: extracellular glucoamylase; mucin-like protein MUC1; protein YIR019c

C:Species: *Saccharomyces cerevisiae*

C&gt;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004

C:Accession: S48478; A26877; B26877; S27281; JC6123

R:Rowley K.

submitted to the EMBL Data Library, October 1994

A:Reference number: S48478

A:Accession: S48478

A:Molecule type: DNA

A:Residues: 1-1367 &lt;ROW&gt;

A:Cross-references: UNIPROT:P08640; GB:Z47047; EMBL:Z38061; NID:G603997; PID:G763364; GS

J.Yamashita, I.; Nakamura, M.; Fukui, S.

J. Bacteriol. 169, 2142-2149, 1987

A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.

A:Reference number: A91831; MUID:87194600; PMID:3106330

A:Accession: A26877

A:Molecule type: DNA

A;Residues: 1-242 <YAM>  
A;Cross-references: EMBL:M16164; NID:gl172522; PIDN:AAA35014.1; PID:gl172525  
A;Accession: B26877  
A;Molecule type: DNA  
A;Residues: 762-1331 <YA2>  
A;Cross-references: EMBL:M16165; NID:gl172523; PIDN:AAA35015.1; PID:gl172526  
R;Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
FEBS Lett. 239, 179-184, 1988  
A;Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae  
A;Reference number: S27281; MUID:89031230; PMID:3141213  
A;Accession: S27281  
A;Molecule type: DNA  
A;Residues: 1-31 <PAR>  
A;Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552  
R;Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
A;Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal growth of *Candida albicans*  
A;Reference number: JC6123; MUID:96323237; PMID:8710886  
A;Accession: JC6123  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-1367 <LAM>  
A;Cross-references: GB:U30626; NID:gl1304386; PIDN:AAC49609.1; PID:gl1304387  
C;Genetics:  
A;Gene: SGD:MUC1; STA2; MAL5; DEX2; SGD:S0001458  
A;Cross-references: MIPS:Y1R019C; SGD:S0001458  
A;Map position: 9R  
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
C;Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
F;5-21/Domain: transmembrane #status predicted <TM1>  
F;1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 16.8%; Score 892.3; DB 1; Length 1367;  
Best Local Similarity 25.6%; Pred. No. 3.6e-11;  
Matches 333; Conservative 107; Mismatches 442; Indels 417; Gaps 65;

QY 9 YLL-----LLLSVFVIOQVSSQELSCKGRG-----FE----- 35  
DB 118 YLDNPTDFATFEVATQDVNS-----CQWMPNFQIQEYLOGSAAQVASSWQWG 169  
QY 36 --SFERGREG-----DCDAQCKYKCCPDYEFSCAECVKNKNNR 73  
DB 170 TTSFDLSTGCNNYDQGHSTQDFPGFYWNIDCDNNC-----GQTKSS 211  
QY 74 TKKKPTKPPVDEAGSLDNGDKFTPTDSTTQHNKVTSPKLTAKINP----- 126  
DB 212 TTTSSSTSSSTTTSSSTSS--SSSTTSSTSSSTTSSTSSSTSSSTTAPATPTTSC 269  
QY 127 -RSLPNSDTSKETSLTVNKETTTNKTQSTGKKTTSKAKETQSTKTSK 185  
DB 270 EKPTPTTCTCKEPTPPHHDTTCTCKKTTTSK-TCT---KKTTTPVTPS--SSTES 324  
QY 186 DLAPTSKVLAKPTPKABTT--TKGPALTTPKE-----PTPTPKPASTTPKEPTPT 237  
DB 325 SSAPV-----PTPSSSTSSSAPVTSSTESSAPVPTPSSSTSSSAPVTSSTES 378  
QY 238 KSAP-----TTPKPAPTTKSAPTTKPEAP-----TTTKEAP---T 273  
DB 379 SSAPVTSSTESSAPVPTPSSSTSSSAPVTSSTESSAPVTSSTESSAPVTSST 438  
QY 274 TPKEAP-----TTTKEAPTTKAPTTKPEAPTTKPKAPTTKPEAPTTKPEPT 327  
DB 439 TESSAPVTSSTESSAPVPTPSSSTSSSAPV-----SSTESSAPV---PTPSS 490  
QY 328 -TTPKPAPTTKE-----PAP-----TPPKPAPTPAKKAPTTTPKPEAPTT 368  
DB 491 STTSSSAPVTSSTESSAPVPTPSSSTSSSAP-APTSSSTSSSAPVTSSTES 549  
QY 369 PKPAPT-----TTKESPTTPKPEAPTTKTSAPTTKPEAPTTTKSAPTT 414  
DB 550 SSAPVPTPSSSTSSSTSSSTSSSAPVPTPSSSTSSSAPVPTPSSSTSSS 609  
QY 415 --PKPSPSTTTKPEAP-----TPKPEAPT-TPKKPAPTTPKPEAPTTPKPEAPT 467

DB 610 PAPTSSSTSSSAPVTSSTSSSAPVPTPSSSTSSSAPVPT-----PSSSTSSSS 665  
QY 468 PTAPKEPAPTTPKETAPTTPKLTTPTPKLAAPTTPKAPTPPELAPTTPEEPTP 524  
DB 666 APVPTPSSSTSSSAPVPT-----SSTESSAPV-----SSTESSAPV-----PTPSSS 713  
QY 525 TTPKEPAPT-TPKAAAPNTPKEPAP-----TTPKEPAPTTPKPEAPTTPKETAPT-TPKG 577  
DB 714 TTESAPVPTPSSSTSSSAPVPTPSSSTSSSAPV-----SSTESSAPVPTPSS 769  
QY 578 TAPTTLKEPAPTTPKKAPKAPKELAPTTTKEPTSTSDK-----PAPTTPKGAAPTTPKEAPT 634  
DB 770 STTESAPVPTPSSSTSSSAPVPT-----PSSSTSSSAPVPT-TPSSSNITSSAPSS 826  
QY 635 -----TPKEPAPTTPKGAP-----TTLKEPAPTTPKKAPKAPKELAPTTTKGP-- 676  
DB 827 PFSSSTESSSAPVPTPSSSTSSSAPVSSSTESSAPV-----PTPSSSNITSSAPSS 882  
QY 677 ---TSTTSKAPAPTTPKETAPTTPKAPAPTTPPET---PPPTTSEVSTPTTKE 731  
DB 883 ITPSSTTESFSTGTT---VTPSSSKYPGSQTETSSTTETTTIVPTKTTTSVTPSTTTI 939  
QY 732 PTTIHKSPDESPELSAETTPKALENSKPEGVPTTKTPAATKPEMTT-----TAKDK 784  
DB 940 TTVVCSGTNSAGETSGCSPKTVTT-----VPTTTTTSVTSSTTTTITVCSGTGNS 994  
QY 785 TTRERLRTTPET---TAAAPKMTKETAATTEKTTESKITATTTQVTS---TTQDTPPK-- 838  
DB 995 AGETTSGCSPKTTTTPVCSSTPSETASSTTTPPTVTTVSTTVVTEYSTKPKGG 1054  
QY 839 -----ITLTKTTTLA--PKVTT-----TKKTTTTEI---MKNPET---AKPK-- 874  
DB 1055 EITTFVTKNIPITYLTITAPTSTVITVNTFTTTTITVCSGTNSAGETSGCSPKTV 1114  
QY 875 ---TTVPCSTGRGEYTTTATTLVTTAVTTVVTTTSTGTSAGKTTTGTGTTKSVPTTVVT 1174  
DB 892 KAPKP-----TSTKKPKTM-----PRVRKPK 913  
DB 1175 LAPSAPVTPATNAVPTTITTECSAATNAAGETTVCSAKTIIVSSASAGENTAPSATPV 1234  
QY 914 TTPTRKMTSTMPELN-----PTSRIAEAM-----IQTTRPN 946  
DB 1235 TTAIPITVITTESSVGTNSAGETTTGTTKSIPTVITLIPGNGAKNVTATATNP- 1293  
QY 947 QTNKSLVEVNPKESDAGGAGETPHMLLRPHVFMPEVT 985  
DB 1294 -----ISIKTTSQATTASASS-----VAPVVTSPSLT 1321

## RESULT 9

S49915  
extensin-like protein - maize  
C;Species: Zea mays (maize)  
C;Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004  
C;Accession: S49915  
R;Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.  
submitted to the EMBL Data Library, June 1994  
A;Description: Pex genes: pollen-specific genes with extensin-like domains.  
A;Reference number: S49915  
A;Accession: S49915  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1188 <RUB>  
A;Cross-references: UNIPROT:Q41805; EMBL:Z34465; NID:9600117; PIDN:CAA84230.1; PID:9600

Query Match 16.3%; Score 874.6; DB 2; Length 1188;  
Best Local Similarity 22.4%; Pred. No. 6.7e-11;  
Matches 259; Conservative 80; Mismatches 394; Indels 424; Gaps 44;  
QY 6 LPVILLILLSVFVI-----QQVSS-----QELSKGRCTEFSEPERGECDCD--- 46



QY	201	ABTTTKGPALTT	-----PKBPPTPTTKPEPA	225
Db	1192	DESTVDGASILTIDTATYYXSEVNHLLTISVVAENTLGAEBTGAQLTIEPKKESVVVVEKQDL	1251	
QY	226	ST	-----TPKEPTPTTKSAPPTPKBPAPTTT	253
Db	1252	SSSEVOKEIAQVKASPEATTITIMETSLTSTKTTTMTGTTTETVSTVGGVTVETKESSE	1311	
QY	254	SAPT	-----TPKEBPAPTTTKEPAP	277
Db	1312	SATTVLGGSGGVTEGSI SVSKIEVVSKTDSQTDVREGTPKRRVSFABEELPKEVIDSDR	1371	
QY	278	-----PAPTTTKEPAPTTTTSKAPTTTPKEPAPTTPKKAPPTTPKEBPAPTTPKBPAPTTTPPKBP	332	
Db	1372	KKKKSPP-DKKEKSPKTEEXPASPTKKTGEBVKSP--KXKSPASPTPKKSPAAAEV	1427	
QY	333	PAPTTKEPAPTP--KEPARTAPKKAPITTPKEPAPTTTPKEPAPTTTTPKSPPTP	385	
Db	1428	KGPTTKKEKSPSPTKKEKSPSPTKKTGDEVKEKSP--PKSP--TKKESKPEKPEDVKS	1482	
QY	386	-----KEPAP-----TTTKSAPTT-----TKEPAPTTTKSAPTTT	415	
Db	1483	PVKKEKSPDATNIVEVSESTIEKTETWTMTMTHESBSRTSVXKKEKTPKVDKPKSP	1542	
QY	416	-----KEPSPTTTPKEPAPTTTPKEPAP--TTPKKAPAPTTTPKEPAPTTTPKEPAPTTTK	465	
Db	1543	TKKDKSPEKSITBEIKSPVKKESPKVBEPKSPATPKKESPKBPASP--TKKSENEV	1598	
QY	466	PAPT	-----APKEPAPITTPKETAPTTPKLTPTTPKLPAPTTPEKPA	507
Db	1599	KSPTKKEKSPKSVVELKSPKPKSP--EKADDPK--SPTKKPK--SPEKSATED	1648	
QY	508	-----PTTPEELAPTTPEEPTTTPPEEAPT--TPKAAAPNTPKBPAPTTTPE--PAPTTPKE	561	
Db	1649	VKSPTTKKESPEKVBEEKPISPTTKESSPTKTDDVKSPTKKESKQTVEEKSPASPTKKE	1708	
QY	562	PAP-----TTPKETAPTPKGTAPTTLKEPAPTTTPKPKAPKELAPTTT	604	
Db	1709	KSPEKSVVBKSPKPKSP-----EKAEBKPKSPTTKKESPEKSAEAEVKSET	1756	
QY	605	-----KEPTSTSDKEPAPTPKGTAPT-----TPKEPAPITTPKEPAPTTTPKGTAPT	651	
Db	1757	KKKSPKESKAEKPKSPTKKESSEPVKADDEVKSPTKKESKPEKVEE--KEASP	1809	
QY	652	LKEPAPTPPKKAPKELAPTTTK-----GPTSTTSD-----KEPAPTTPKETAP	694	
Db	1810	KKE---KTPKSAABELKSPTKKESKSPSTKKTGDSKESKPEKPEKSPTPPKKSP	1866	
QY	695	TTPKPAPTTTPKAPPTTPPTTPPTTSEVSTPT-----	727	
Db	1867	GSPKKKKSPEAKPKPAPK-----LTRDLKLQTVNKTDLAHFEVVVVEHATECKWFLDGKE	1922	
QY	728	TTKEPTTIHKSPD-----ESTPELSAEPKPKALENS	758	
Db	1923	ITTAQGVTVSKDQOFEPFCSIDTTMFGSGTVSVVSNAGSVEYTKELKVLPTK	1977	
QY	759	PKBPGVPTTKTPAATKPEMTTAKD-----KITERD-----	789	
Db	1978	-----ETK-----KPEFTDKLDEMVTGKDTQCMDIALHSLPKWYQNGNLLEDGKN	2025	
QY	790	-----LRTTPETTTA-----APKMT--	804	
Db	2026	GVTIKNEENKSSLIIPNAQDSGKITIVASNEVSGSESSAQTVNPPPTPIVVDGPKSVT	2085	
QY	805	KETAT-----TTEKTESKITATTTQVTTSTTQDTPPKIT-----TL	842	
Db	2086	IKETETAFKATISGFPAPTVKVTINEKIVESRTIITIKTEDVYTLKISNAKIEQGTV	2145	
QY	843	KTT-----TLAPKVVTTTKKIT	859	
Db	2146	KVTAQNSAGDSKQADLKVEPNVAKPF--KSQLTDKVADEGBEPLRWNLNLELDGSPSGTE	2202	
QY	860	-----TTEIMN-----KPEE-----TAPKQRA--TNSKAT	884	

Db	2203	VSWLLNGQLT	KS	DI	VQ	VD	H	GD	GT	VH	VI	TA	EA	CP	EM	S	GL	TA	KA	NA	AG	CE	TS	SA	K	V	T	2362	
Qy	885	PKPQKPT	KA	PK	K	P	T	S	K	P	T	S	K	P	T	S	K	P	T	S	K	P	T	S	K	P	T	S	903
Db	2263	NGGNKKP	EF	VQ	PA	QN	HE	TT	LE	BS	VK	FS	AI	VT	GK	PM	PN	VT	WY	LN	NK	KL	IQ	SE	EV	KV	VH	2322	
Qy	904	KTMP	RV	RP	K	P	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	925		
Db	2323	TGKTS	IR	IQ	KP	LE	HN	GT	IR	VE	AE	NV	SG	VK	QV	TA	QK	LV	P	---	---	---	---	---	---	---	2381		
Qy	926	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	961		
Db	2382	EDVK	TA	NV	EG	PE	PS	VA	ML	NG	EP	VS	---	---	---	---	---	---	---	---	---	---	---	---	---	---	2431		
Qy	962	DAGG	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	982		
Db	2432	EQAG	EL	CE	BA	TNP	VG	SK	RD	VQ	LA	VK	KG	DA	PT	FA	KN	LE	DR	LI	TE	GE	BL	TL	MD	AK	LN	2491	
Qy	983	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Db	2492	KPKIT	WL	KD	GE	IT	S	D	GH	Y	2510	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		

RESULT 11

T34513

hypothetical protein ZK783.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T34513

R;Favella, A.; Vaudin, M.

submitted to the EMBL Data Library, August 1994

A;Description: The sequence of C. elegans cosmid ZK783.

A;Reference number: Z21536

A;Accession: T34513

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-3507 <FAV>

A;Cross-references: UNIPROT:Q23587; EMBL:U13646; PIDN:AAC24418.1; GSFDB:GN00021

A;Experimental source: strain Bristol N2; clone ZK783

C;Genetics:

A;Gene: CRSP:ZK783.1

A;Map position: 3

A;Introns: 14/1; 48/2; 84/1;

```

QY 111 -----KVSTPKITTAKEINPR-----PSLPPNS- 134
Db 466 ISSHGAFTTTGEQIIEGSGIQTSSGGSITVTRGLIPKQVELITTSRGLACTSYCPFNSE 525
QY 135 -----
Db 526 CVGGYCEVSGYGNALVGCBDIDECITEICNIEANWCVNLIIGFVCCNPTNATHDDCI 585
QY 135 -----DTKSTSLTV---NKETTIVETKETTINKOTS----- 163
Db 586 DELTKVKIYAMIIFLLKGLIEITKBLHWIGNEEDIV---VATRSNHSTSOQLITQV 642
QY 164 -----TDGK-----EKTSAKETQSIKTSKADLAPTSKVLAKPT---P 199
Db 643 VQOSRNFSTGQIILTRGVSSGEAVTQTTDADEFL--EISAADLAGSGSITLPTLLEP 700
QY 200 KAE-----TTTKGPALT-TPK----- 214
Db 701 KIEGSGKASGVWTEDEGEDEDLMBEGSGSWTTINGTIGTSPRSEGTIRVRITLIG 760
QY 215 ---EP-TPTTTP-----KE----- 223
Db 761 EDGEPEATKPGISAPDKTGGSKTESDGEKLTVEKDGEAKQSGGSATSSGKKSEATS 820
QY 224 -----
Db 821 GSSSSSAKSGTGFASGSGASGSSGSGVSGESSVSTESGSGFTSSSGSVGSEATGST 880
QY 224 -----PASTTPKEPTTT-----IKSAPTTKPEAPATTK----- 253
Db 881 GVDGSSGKPSKSTBEKLPFTKNGEKSPISGDDTIGKESSEETTSRKPIEGSDSLTEGS 940
QY 254 -----SAPTTKPE-APTIT----- 267
Db 941 GSSEWFTGSKGHFESGSKVSVTSGKPTQSGAEGSGGPKVKGPGAPBITTDGBESS 1000
QY 268 -----KEPAPTTKPEAPTTT----- 283
Db 1001 TSTGDKSGGKPADKSDNKNVPGTKDNPOITTDGEDSTSETSGEGQPKGSKQPPGDK 1060
QY 284 ---KEPA-----PT-----TTKSA----- 294
Db 1061 GSEVKKTEVDGPGNLGSGTKGKSNVPLKPTDLPERGSGILTTSSGKNSFTBHGTKLER 1120
QY 295 -PTTKPEAPTTTP-----KKPAP-----TTP----- 314
Db 1121 LPPKTEKSSFPOLGLBISAGKKPPEPGTKEVGLLEILWESTTPGSTLLDSDSVGLEI 1180
QY 315 -----KEP-----APT 320
Db 1181 SGSDLTKATKPHVEIEGSGTGDEEITATTRDVSKSTKPRVEVDGDNGETSGVDGKPT 1240
QY 321 TPKEPTPTTKPEPA---PTTKPEAPT-----TPKEP----- 348
Db 1241 TP-APTSSASBSTSRIPITTEASPEGSGGAGVPESPDGSGESSTAPDGVSPITSSAT 1299
QY 349 APTAPKAPPTTP---KEPAPTTK---EPAPTTTKPESPPTTKPEPA---PT--- 391
Db 1300 APEVPTSSASTPDAVEESGIFSTSKPTAEPLLETTAPSTEVTSPEGSGTEESTLPTEGS 1359
QY 392 ---TKSAPTTTKPEPA-----PTTKKSAP----- 412
Db 1360 GESITSSAPTIV--EPATVLPQNRNEKPEPTKDTFALPTTTTGCAQANDSSVENTKCTSS 1417
QY 413 -----TTTKPEP----- 418
Db 1418 DECGLDALCERRTGVCRCEPFGEPAGPKKSCVDVDECATGDHNCHEARCONVVGAYACF 1477
QY 419 SPT-----TTK 424
Db 1478 CPTGFRKADGSCQDIDECTEHNSTCCGANAKVKNKPGTYSCECENGFLDGYQCVPTTK 1537

```

```

QY 425 EPAPTTPK-----EPAP----- 436
Db 1538 KPCDSTOSSKSHCSSESNNSCVDTVDSVECKECMGGYKKGKVCEDINECVAEKAPCSL 1597
QY 437 -----
Db 1598 NANCVMNNGTFCSCQKQYRGDGFMCCTDINECDERHPCHPAECTNLLEGSKCECHSGFE 1657
QY 437 -----TTP----- 439
Db 1658 GDGIKKCTNPLERSCEDVEKFCGRVDHVSLSVRIYNGSLSSVCECBPGPFPEKESNCV 1717
QY 440 ----- 439
Db 1718 DIDECEBSRNNCDPASVCVNVNTEGSRCECAEGYEGGVCTDIDECDRGMAGCDSWAMC 1777
QY 440 ----- 439
Db 1778 INRMGSCGCKMAGYTGDGATCIKIEEPKSDKTACTDEWSRLCELEKKQCTVDEBEVPQ 1837
QY 440 ----- 439
Db 1838 CGACLPGHPIPINGTQCSLQISGLCAQKNDCKHAECIDIHPDSHFCSGPDGFIGDGMICD 1897
QY 440 ----- 439
Db 1898 DVDECNNAGCDDENTKCENTIGSPNCVCLGFKVDEKCVDEKQPNREKIEIDENS 1957
QY 440 ---KKP-----APTTKPEAPTT-----PK 456
Db 1958 SSSNSGOEKTTKGIYSSTSATSESTAEPHVTTSSISSTTKDMTSSKSPENVMTSSE 2017
QY 457 EPAPTTTKPAPTAPEKAPPTPKET---APTTPKKLT-----PTTPE--- 496
Db 2018 SPEVSTSSSKSTTASETTVSTESSESSSEAPLTSPPATTTVEVITESSVSKTTPKESSS 2077
QY 497 ---KLAPTTPE---KPAPTTPELAPTTPEEPTPTTPEPAPTTPKAAAPNTPKBP 546
Db 2078 EITVLUSSKSPEVTESSVSSPSTPS---TTSQSVTSTVPETSKSTLSSEAPVITSTSP 2133
QY 547 ---APTTPKPEAPTTKPEAPTTPKET---APTTPKGT---APTTLKEAPTT 590
Db 2134 TEVHTSSEIKPSLSASITTDNTSTPSSSLASVKSSTAPEGTSASVAPVLUSSLPDV 2193
QY 591 PKKAPKELAPT-----TTKEP-----TSTTSKAPPTPKGTAPT 627
Db 2194 -SQPSTKTFDATESSTVQASSETSSGTSVKSTSEPESHVTKLSITSSNPSSVVPVSPKST 2252
QY 628 PKEPAPTTKPEAPTTPKG-----TAPTTLKE 654
Db 2253 PTVPEST--EQPTSTTPGQSLTPMNSNSEVLTTSPEHVLSSSLSPDVSSQSTTPNNLSE 2310
QY 655 PAPT-TPKKAPKEL---APTTLTKGT-----STT-----S 681
Db 2311 SSTVETPKTSSEVLSNSEPSTTEAPTLSPDLISITNNLSSQSTVSTEDSEISSENS 2370
QY 682 DKP-----APTTP-----KETAPTTP----- 697
Db 2371 EKPTSAPELVTSVTHVASSPDVPTESSEPDDLGSSTENIPEASSKQTLISSTPTPDT 2430
QY 698 ---KEPAPTTKPP-----APTPE----- 714
Db 2431 TASEEPTKSTGMSPLSTTSNVLSSESTTPRESSKSPVSSSTEGISVVVTSTEFKVPST 2490
QY 715 -----TPPPTTSEVSTPTTTKEPT---IHK--SPDESTEPLSAEP 751
Db 2491 ISSVLEEDLTKTTPSPILEETTTASETSEPLTDSLTVSVRIHELTTSSENVPKSEST 2550
QY 752 PKALENSP-KEPG-----VPTT-----KTPAATKPEM--- 777
Db 2551 TSSESSKPSQBPAGILTSTVVVPTSSVSLITASEIEALTSNTPFKQGTPTTTSFKSLVK 2610
QY 778 ---TTTAKOTTERDLRTPPTTTAAPKMTKETATT-----TEKTTESKIT 820

```

2611	STTSPSTVTSSEPSSTKRTTVSTTVSTTTPTTEETTSLSLTLTAAPSKPTSTSTESSA	2670	QY	198	-----TPKAETTTKGPALATPKPTPTT---	PKEPASTTTPKEPTPTTIKSAPTT	243
821	ATTIQVTS-----TTTQDTTPFKITLTKT-----TLAPKVTTTKK	856	Db	192	BQTSNDNTEIASPMETNITTEATTSVSPSVSLASEDETTVTAABSTTVIAEVSIT	251	
2671	PTTPAKTSKPSNVSSSTRKSTENVETSTSQSGLSSTMSSTSEPETNAPAVTSSE	2730	QY	244	PKBP-----APTTSKAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT	300	
857	TITTT-----EI	863	Db	252	TBEPPTTABSSTTKKST---KAPATTEPTPTTTEE---VTTEAETSTTSSETSEK-	304	
2731	ASSTTLEENSSSTSPSSBASVKLSLFPESITSEAVTVSSRAPABITSSSHREISTV	2790	QY	301	PAPPTP---KKPAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTP	351	
864	MNKPBP-----TAKPKD-----	875	Db	305	-----PTPLIUNKIAGAPAGK---PETTHFPVTGTP-----NPDTEATEPFFVAKSEDKMT	353	
2791	SEEPSPEIPLSTTVSPNVVTVASSIPSEBILSSVTSSTPRVLITGTPDDLLIVSVTP	2850	QY	352	APKAPAPTTTPKEPAPTTTPKEPAPTTTKE-----PSPTTPKEPAPTTTPKSAP-----TT	399	
876	-----RATNSKA-----TTPKPQKPTKAPKPTSTKPKTPRVRKPK	913	Db	354	LKATAAETTTQQTTEVT-DGPEKETTKNWSIEPIITTVPLVETSTTSTASKESDGHHT	412	
2851	SHGNRRQNIWASSVPNSNSTPIILPSESLTTPQPPPTTTTAKPATISGKRGPPSIQPPA	2910	QY	400	TKBPAPT-----TTKSAPPTTPKEPSPTTKE---PAPT-----TPK-----EPAPPTP	439	
914	TTPTP-----RKMTSTMPENLPTSRIAEAMLOT-----	941	Db	413	LKLKVTTADSDSTESATTVKPFNEETTTKSHVVPKTKGTVKVPKLELSFDEPTEIT-	471	
2911	EMFTTAPPPPSNGYGEETNQBEQVIST-----TTTEAPSLCSTVTCHSLATCEQST	2964	QY	440	KKAP-----	444	
942	-----TTRPNQTPNSKLVNPNKSEBAGAGET-----P	971	Db	472	KAPHPKGLLEKTYHFLVSDNFARYSEAKENDDYNHLDYHNYREAKEPTTTESSITTEEV	531	
2965	GVICIRDRGFTGDTTACKSKSTADCSLPSLCADKAKCDNSTRSCEDAGYIGDGYVCS	3024	QY	445	TTKEPA-----PTTPKEPAPTTTKKAPATPAPKEPA---PTTPKETAPTTPKLTPTTPEK	497	
972	H---MLLRPHVFWPE	983	Db	532	TTTEEPANTGNPPTTEN---PTTTEQPTSTAESTTTALPFTTEQTV-----TTEE-	578	
3025	HPQDCVLRLNLCSP	3039	QY	498	LAPTTPEKAPPTTPEELAPTTPEEPTPTTPEEAPPTPKAAAPNTPKAPPTTPKEPAPT	557	
RESULT 12			Db	579	--PTTAKSTATQ---KPTTQESVST---EKSTTKKA---STTEE---PTTDEPTT	624	
T25697	hypothetical protein F16F9.2 - Caenorhabditis elegans		QY	558	TPKEPAPTTPKETAP---TTPKGTAPTTLKEPAPTTPKAPKAPKELAPTTTKEPTSTTSDK	614	
C;Species: Caenorhabditis elegans			Db	625	T---BSSTTKGATTPBELSTTSEETTELK---ITTE-----GSTTTEEPITTAIPA	670	
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			QY	615	PAP-----TTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTLKEPAPTTP---	660	
C;Accession: T25697			Db	671	EASTGIITDEETSTTSTTPEITSTKE--IVTESAITQTSVSVSVESSTPQLPWRKAI	728	
A;Submitted to the EMBL Data Library, August 1996			QY	661	-----KKAPKAPKELAPTTKGP-----TSTTSDKAPPTTPKE	691	
A;Description: The sequence of C. elegans cosmid F16F9.			Db	729	VNKFKHNLVLEKRLKKESTSTGSDSSETTTVVAENIDVTTIEKEKVQVQTPIT	788	
A;Reference number: Z20071			QY	692	TAPTTTPKEPAPTTPKAPPTTPE---TPPPTTSE-VSTPTTTPKEPTTIHKSPDESTBELSA	748	
A;Accession: T25697			Db	789	TEKSTTQETTTTTEKTTSTKTTTEKPTTSESATTTETSEPT---TEST---	838	
A;Molecule type: DNA			QY	749	EPTPKALENSPKBGPVPTTKTPAAKPEMTTAKDKTTERDLRTPPETTTA-----	799	
A;Residues: 1-1229 <FUL>			Db	839	-----TVDTSATTESSATAAETTTTSAE---TSETTTSESAAFTGE	878	
A;Cross-references: UNIPROT:Q94185; EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9			QY	800	-----APKMTKETATTTTEKTTESKITATTTQ-VTS	828	
A;Experimental source: strain Bristol N2; clone F16F9			Db	879	SPENTALQSSSQKSEBENESSAABKPGARRDFVPKHKTTPKPAETTSVAASATTTTEPTT	938	
C;Genetics:			QY	829	T---TTQDTTPFKITLTKTTLAPKVT---TTKKTITTTTEIMNKPEETAKPKDRATNSKA	882	
A;Gene: CESP:F16F9.2			Db	939	TEKSTTLETPIEATTLNEVTPGPAVTCAPVDETTINTLELLSK-----INNTOI	988	
A;Map position: X			QY	883	TTPKPQ-----KPTKAPKPTSTKPKTPRVRKPKTTTPPKMTSTM	925	
A;Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3			Db	989	SQPKPTDISKTDALSSLLISGLIGSFTKAPMPTI-----HTTDAAFVTA	1035	
Query Match	15.7%;	Score 844.3;	Db 2;	926	PELN-----PTSRIAEAMLOTTPRNPOTPNKSLVENVNPKSDAGAGETPHM	973	
Best Local Similarity	26.0%;	Pred. No. 2.8e-10;		Db	1036	ASLNDGSDKKIIDBAQPTDEIRRA-----QPTN-----	1063
Matches	334;	Conservative 95;	Mismatches 360;	QY	974	LLRPHVFMPEVTPDMVDLPRVFNQGI	999
QY	2	AWKTLPIYLLLLSVFVIQVSSQELSCGK-RCFESFERGRCDCDAQCKYDKCCPDYE	60	Db	1064	-----EMDKEMEPEKRIEORI	1080
Db	4	AW-VVSFAFLILGNVQSSLSLKTINSGSRDPKIVKHKKNTCTCSC---KCVDP--	56	QY			
QY	61	SFCAEVKDNKQRTKKTTPKPPVVDGAGSLNDGDFKVTPTDSTT---QHNKVSTSP	116	Db			
Db	57	-----APSNPF-----DVSTTSSINNNNDVIGP	80	QY			
QY	117	KITTAKPINRPSLPNPDTSKETSITLVNKETT-----VETKETTITNKQTS--	163	Db			
Db	81	S-----GDSNFTGSSWFQIEATVGGQTVKSEHNIDSSVEVEKKVTTSTDASTNA	131	QY			
QY	164	-TDGKEKTTSA-----KETQSIEKTS-----AKDLAPTSKVLAKP	197	QY			
Db	132	PTTGKDSITTEIITGIWVINSKSESVTMSITRFTSTLTSPTTELLTSPETLVSTDSSTST	191	QY			



[illegible]

```

Query Match      15.3%; Score 822.5; DB 1; Length 1344;
Best Local Similarity 23.3%; Pred. No. 8.8e-10;
Matches 318; Conservative 80; Mismatches 370; Indels 595; Gaps 68;

7 PYILLLSVFIQQVSSQELSCKRGCFESFERGECDCDACKKYDKCCPDYEFCAEV 66
   | : ||||| :
8 PFFILLLLTLVT----- 19

```

```

Db 972 -----DTRP-----APGSTAPPAGHGVTSAPDTRPA 996
QY 887 FQKPTKAPKPTSTKKPTMPRVKPKTTPRKMTSMBELNP----- 930
Db 997 P--GSTAPPAGHGVTSAPDTRP-----APGSTAPPAGHGVTSAPDNRPALGSTAPPVHNVTSAS 1051
QY 931 --TSRIAMLO--TTTPNPTPSK-----LVEVNPKSEDAGGAGET 970
Db 1052 GSASGSASTLVHNGTSARATTPASKTPPSIPSHSHSDTPTTLASHSTKTDASSHTHTV 1111
QY 971 PHMLLRHVFVEVT-----PMDY 990
Db 1112 PPLTSSNHSTSPQLSTGVSFPLSHISLNQFNSSLEDPSDY 1154

RESULT 15
B48666
cell proliferation antigen Ki-67, short form - human
C:Species: Homo sapiens (man)
C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Mar-2004
C:Accession: B48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerdes
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiqui-
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: B48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 -SCH-
A:Cross-references: EMBL:X65551
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F;29-91/Domain: kinase interaction domain homology <KIH>

Query Match 15.1%; Score 811.2; DB 2; Length 2897;
Best Local Similarity 15.6%; Pred. No. 4e-09;
Matches 383; Conservative 128; Mismatches 404; Indels 1548; Gaps 92;

QY 22 VSSQELSK-----GRCFSEFGRGCCD 46
Db 382 ISNQKWFREDLSGIAEMFKTPVKEQPQLTSTCHIALNSENLLGQFQGTDSGE----- 437
QY 47 AQCKYDKCCPDYESCAEVKDNKQRTK---KPTPKPV---VDEAGSGLDNGD--- 96
Db 438 -----PLUPTSESGNVFFSAQNAAKQPSDKCASPLRRQCIRE-----NGNVAK 484
QY 97 -----FKVT---TPDTSIQHNKYST---SPKITAKPINRP----- 128
Db 485 TPRTNYMTSLTKTSDTETEPSKTVSVNRSRSTEFNRNIQKLPVESKSEETNTEIVEC 544
QY 129 ----- 128
Db 545 ILKRGQKATLLQORREGMEIERPFETYKENIELKENDEKMKAMKRSRTWGQKCAPMSD 604
QY 129 -----SLP-----PNSDTSKETSITVKN-----ETTVETK- 153
Db 605 LTDLSLPDTMLMOTARGQNLQOTDHAKAPKSEKGIKTMPCQSLQPEPINTPHITQ 664
QY 154 -----ETTTTNKQISTDGE----- 168
Db 665 QLKASLGKGVKBEALLAVGKFTRTSGTTHTRREPAGDGKSIKRTFKESPKQILDPAARVT 724
QY 169 -----KTSKAKETQSIETKSAKDLAPTSKVLAKPTKAE----- 202
Db 725 GKKWPRTPKEAQSLE-----DLAGFKELFQTPGSEESMTDEKTTKIACKSPPPESVD 779
QY 203 --TTTK-----GPALTTPKEP-----TPT--- 219
Db 780 TPTSTQMPKRSIRKADVEEFLALRKLTPSAGKAMLTPKPAGDEKDIKAFWTFVQKL 839
QY 220 -----TPKEPAST-----TPKEP----- 232
Db 840 DLAGTLPKSRLOTPKEKAQALDLAGFKELFQTPGHTTEELVAAGKTTKIPCDSPQSDP 899

```

```

QY 233 --TPTTIKSAFT-----TPKEPAP----- 249
Db 900 VDTPTSTQRPKRSIRKADVEGELLACRLNLPMSAGKAMHTPK---PSVGEEKDIIIFVGT 956
QY 250 -----TTTKSAFTTPKEPAPT-----TTKEPAPT 273
Db 957 PVQKLDLTENITGSKRRPQTPKEBAQALDITGFKELFQTPGHTTEEVAAGKTTKMPCES 1016
QY 274 TPKEPAPTTPKEPAPTTPKSAFTTPKE----- 300
Db 1017 SPPEADTI-----PTSTRQPKTPLEKRDVQKLSALKLTQTSGETTHYDKVPGBDX 1070
QY 301 -----PAPTTPKAPAPTTPKEPA-----PTTPK- 323
Db 1071 SINAFRETAKQKLDPAASVTGSKRHPKT-KEKAQPLEDLAGWKELFQTPVCTDPTTHEK 1129
QY 324 -----EPTPT-TPKEPAPTKE-----PAPT----- 343
Db 1130 TTKIACRSQDDPVDPTTSSKPSQSKRSRKVDVEEFFALRKRTPSAGKAMHTPKPAVSGE 1189
QY 344 -----TPKEPA----- 349
Db 1190 KNIYAFMGTPVQKLDLTENLTGSKRRLQTPKEKAQALDLAGFKELFQTRGHTTEESMTND 1249
QY 350 -----PTAPKKA-----PTTPKEPAPT- 367
Db 1250 KTAACKSSQPDLDKNPASSKRRLKTSGLGKGVKEALLAVGKLTQTSGETTHHTHTPTG 1309
QY 368 -----TPKEPA-----PTTKKPSPP 382
Db 1310 DGSKMAFMESPQILDSASLTGSKRQLTPPKGSEVPEDLAGFIELFQTPSHTK- SM 1368
QY 383 TTPK-----EP-----APTTPKSA----- 397
Db 1369 TNEKTTIKSVYRASQPDVLTPTSSKQPQKRSRKADTEEFRAFQKTPSAGKAMHTPKP 1428
QY 398 -----TTTKEPA-----P 405
Db 1429 AVGEKDIINTFLGTPVQKLDQPGNLPGSNRRLQTRKEKAQALBELTGRELFPQTPCTDNP 1488
QY 406 T---TTK---SAPTTPKEPSPITTK-----PA----- 427
Db 1489 TADEKTTIKLKCSPOSDPADTPTNTKQPKRSLKADVEEFLAFKRLTPSAGKAMHTP 1548
QY 428 -----PTTPKEPAPT-----TP--- 439
Db 1549 KAAVGEKDIINTFVGTPVEKLDLGNLPGSKRRPQTPKEKAQALDLAGFKELFQTPGHT 1608
QY 440 -----KKAAPTTPKEPA----- 451
Db 1609 EESMTDDKITEVSCSPQDPVKPTTSSKQRLKISLGKGVKEEVLVPGKLTQTSKITQ 1668
QY 452 -----PTTPKEPA----- 459
Db 1669 THRETAGDGSKIAFKESAKQMLDPANYGTGMERWPRTPKEBAQSLDLAGFKELFQTPD 1728
QY 460 -----PTTKKAPAPTAPKEPAPT---TPPKET---APTTP-----KKLTPTT--- 494
Db 1729 HTEBSTTDDKTTKIACKSPPPESMTPTSTRRRPKTPPLGKRDIIVEELSALKQLTQTHTD 1788
QY 495 -----BEKLAP---TTPEKAPAPTTP-----BELA----- 515
Db 1789 KVPGEDKGINVFRETAKQKLDPAASVTGSKRQPTPKGAQPLEDLAGKELFQTPVCT 1848
QY 516 --PTTPESPTTPPEPAPT----- 533
Db 1849 DKPTTHEKTTIACKSPQDPDPVGTPTIFKPOSKRSRKADVEEESLALRKRTPSVGKAMD 1908
QY 534 TPKAAA-----PNTPKKEPA----- 547
Db 1909 TPKPAGDEKDKAFMGTPVQKLDLPGNLPGSKRWPTPKKAQALDLAGFKELFQTPG 1968

```

Search completed: October 13, 2004, 11:59:33  
Job time : 58.4912 secs

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	7261.9	99.8	1404	2	AA826049	Aar26049 MSF precu
2	7261.9	99.8	1404	4	AA829773	Aab29773 Human meg
3	7261.9	99.8	1404	4	AA860568	Aab60568 Human meg
4	7261.9	99.8	1404	8	Adm98014	Adm98014 Human meg
5	7258.9	99.8	1415	4	AAU32262	AAU32262 Novel hum
6	7232.9	99.8	1404	7	ADK65839	ADK65839 Angiogene
7	7043.7	96.4	1320	7	ADK65819	ADK65819 Angiogene
8	6788.7	93.3	1270	8	ADK67911	ADK67911 Human ext
9	6774.6	93.1	1311	8	ADK67912	ADK67912 Human ext
10	6685.9	91.9	1299	4	AAW24332	AAW24332 Human EST
11	3552	48.8	902	4	AA829778	AA829778 Human MSF
12	2929	40.3	551	4	ABU53253	ABU53253 Human tes
13	2920	40.1	546	4	ABU53252	ABU53252 Human tes
14	2850.9	39.2	538	5	AAO18834	AAO18834 3' cartil
15	2757	37.9	513	4	ABU53254	ABU53254 Human tes
16	1954.5	26.9	452	2	AA880041	AA880041 Human meg
17	1802.7	24.8	472	4	AA860569	AA860569 Bovine MS
18	1401.6	19.3	292	5	AAU11261	AAU11261 Human HSP
19	1371	18.8	519	4	AAW24516	AAW24516 Human C9
20	1371	18.8	5179	6	ABP55365	ABP55365 Human col
21	1371	18.8	5179	6	ABO07258	ABO07258 Human p53
22	1371	18.8	5179	7	AD048091	AD048091 Human pro
23	1371	18.8	5179	7	ADP44998	ADP44998 Human pro
24	1196.3	16.4	8991	6	AAU08487	AAU08487 S. pneumo
25	1190.4	16.4	1664	2	AAW43106	AAW43106 C. thermo

PR 18-JAN-1991; 91US-00643502.  
PR 10-SEP-1991; 91US-00757022.  
XX (GEM) GENETICS INST INC.  
XX Turner K, Clark SC, Jacobs K, Hewick RM, Gesner TG;  
PI  
XX WPI; 1992-284660/34.  
DR N-PSDB; AAQ27223.  
XX  
PT New human mega-karyocyte stimulating factors - for treating immune  
PT deficiencies, cancer, exposure to radiation or drugs, bacterial and viral  
PT infections, etc.  
XX  
PS Claim 1, 2 and 3; Fig 1; 87pp; English.  
XX  
CC The sequence given is a full length translation from the megakaryocyte  
CC stimulating factor (MSF) precursor. The sequence covered by exons II, III  
CC and IV encodes megakaryocyte stimulating factor (MSF). This sequence is  
CC modified by the addition of an N-terminal sequence encoding a secretory  
CC leader, an initiating methionine preceding exon II and a terminating  
CC codon following exon IV. The cDNA sequence given contains sequences  
CC derived from human megakaryocyte colony stimulating factor (meg-CSF).  
CC Exon I contains the initiating methionine, and encodes a classical  
CC mammalian protein secretion signal sequence. The sequence encoding the  
CC original meg-CSF includes exons II-IV and is thought to terminate in the  
CC region between amino acid residues 134 - 147. The primary transcript of  
CC this gene may be cleaved in different ways to yield a family of mRNA's  
CC each encoding a different MSF protein. Exons V and VI are thought to be  
CC related to the activity of the factor and are also implicated in the  
CC stability, folding and processing of the molecule. These exons are also  
CC thought to play a role in the observed synergy of MSF with other  
CC cytokines. Exons V - XII are believed to be implicated in the processing  
CC or folding of the appropriate structure of the resulting factor, i.e. one  
CC or more of these exons may contain sequences which direct proteolytic  
CC cleavage, adhesion, organisation of the cellular matrix or extracellular  
CC matrix processing. Both naturally occurring and non-naturally occurring  
CC MSF's may be characterised by various combinations of alternatively  
CC spliced exons from this sequence, with the exons spliced together in  
CC differing orders to form different members of the MSF family. (Updated on  
CC 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 1404 AA;  
  
Query Match 99.8%; Score 7261.9; DB 2; Length 1404;  
Best Local Similarity 97.1%; Pred. No. 7.3e-177;  
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;  
  
QY 1 MAWKTLPIYLILLLSVFVIQVSSQ----- 25  
DB 1 MAWKTLPIYLILLLSVFVIQVSSQDLSSACRGCGEYSRDATCNDCYNCHYNECCPDF 60  
QY 26 -----ELSCKGRCFESPERGECDDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 79  
DB 61 KRVCIAELSCKGRCFESPERGECDDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120  
QY 80 PPGASQTIKSTTRSPKPPNKKTKVIESEBITTEHSVSENQESSSSSSSSSSSTII 139  
DB 121 PPGASQTIKSTTRSPKPPNKKTKVIESEBITTEHSVSENQESSSSSSSSSSSTII 180  
QY 140 KIKSKNSAANRELQKLVKNDKNRTKKKTPKPPVWDAGSLDNGDFKVTTPDTST 199  
DB 181 KIKSKNSAANRELQKLVKNDKNRTKKKTPKPPVWDAGSLDNGDFKVTTPDTST 240  
QY 200 TQHNKVSPTKITTAKPINRPSLPPNSDTSKETSLTVNKETTVETKETTNNKQSTDG 259  
DB 241 TQHNKVSPTKITTAKPINRPSLPPNSDTSKETSLTVNKETTVETKETTNNKQSTDG 300  
QY 260 KEKTTSAKETQSIEKTSADIAPTSKVLAKETPKAETTTKGPALTTTKEPTPTPKBPAS 319  
DB 301 KEKTTSAKETQSIEKTSADIAPTSKVLAKETPKAETTTKGPALTTTKEPTPTPKBPAS 360  
QY 320 TTPKEPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 379

DB 361 TTPKEPTTTIKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 420  
QY 380 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 439  
DB 421 APTTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 480  
QY 440 EPAPTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 499  
DB 481 EPAPTAPKKPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 540  
QY 500 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 559  
DB 541 TTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 600  
QY 560 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 619  
DB 601 APTAPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 660  
QY 620 PEEPAPTTTCAAAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 679  
DB 661 PEEPAPTTTCAAAPNTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 720  
QY 680 APTTPKKPAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 739  
DB 721 APTTPKKPAPKELAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 780  
QY 740 TAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 799  
DB 781 TAPTTLKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 840  
QY 800 KPAPTTTPEPPTTSEVSTPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 859  
DB 841 KPAPTTTPEPPTTSEVSTPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEP 900  
QY 860 TKTPAATKPBMTTAKDKTERDLRTPPTTAAAPKMTKATTTTTEKTESKITATTTQV 919  
DB 901 TKTPAATKPBMTTAKDKTERDLRTPPTTAAAPKMTKATTTTTEKTESKITATTTQV 960  
QY 920 TSITTTQDTTFFKLTITLTKVTTTLPKVTITTKTITTTTMMKPEETAKPKDRATNSKATTPK 979  
DB 961 TSITTTQDTTFFKLTITLTKVTTTLPKVTITTKTITTTTMMKPEETAKPKDRATNSKATTPK 1020  
QY 980 PQKPTKAPKXPTSTKKPKTMVRVKPKTTTTPRMTSTMPELNPTSRIAEAMLQTTTRPN 1039  
DB 1021 PQKPTKAPKXPTSTKKPKTMVRVKPKTTTTPRMTSTMPELNPTSRIAEAMLQTTTRPN 1080  
QY 1040 QTPNSKLVEVNPXSSEDAGGAEGETPHMLLRPHVFMPEVTPDMOYLPRVPMQGIINPMLS 1099  
DB 1081 QTPNSKLVEVNPXSSEDAGGAEGETPHMLLRPHVFMPEVTPDMOYLPRVPMQGIINPMLS 1140  
QY 1100 DETNINCNGKPDVGLTTLNGLTVAFRGHYFWMLSPPSPSPARRITVWGIPIPSIDTVFT 1159  
DB 1141 DETNINCNGKPDVGLTTLNGLTVAFRGHYFWMLSPPSPSPARRITVWGIPIPSIDTVFT 1200  
QY 1160 RCNCEGKTTFFKDSQYWRFTNDIKDAGYKPKIFKFGGLTGQIIVAAALSTAKYKNWPSVY 1219  
DB 1201 RCNCEGKTTFFKDSQYWRFTNDIKDAGYKPKIFKFGGLTGQIIVAAALSTAKYKNWPSVY 1260  
QY 1220 FFKRGSGIQYIYKQEPVQKCPGRRPALINPVYGEVQVRRRRRFEAIGPSQTHIRIQY 1279  
DB 1261 FFKRGSGIQYIYKQEPVQKCPGRRPALINPVYGEVQVRRRRRFEAIGPSQTHIRIQY 1320  
QY 1280 SPARLAVQDQGVHLNHEVKVSTLWREGLPNVVTSAISLNPDKPDGYVYAFSKQOYYNIDV 1339  
DB 1321 SPARLAVQDQGVHLNHEVKVSTLWREGLPNVVTSAISLNPDKPDGYVYAFSKQOYYNIDV 1380  
QY 1340 PSRTARAITTTRSGQTLKSVWYNCP 1363  
DB 1381 PSRTARAITTTRSGQTLKSVWYNCP 1404



1201	RCNCEGKTFFFKDSQWVRFTNDIKDAGYKPIKFGFGGLTQIIVAAI	STAKYKXWNPESVY	1260
1220	FFKRGGSIOQYIYKQBPVKQCPGRRPALNPVYTGEMTVQRRRFR	FAIGFSQTHIRIQY	1279
1261	FFKRGGSIOQYIYKQBPVKQCPGRRPALNPVYTGEMTVQRRRFR	FAIGFSQTHIRIQY	1320
1280	SPARLAYQDKGVLHNEVKYSILWRGLPNVVTSAISLPNIRKPGDGY	YVYAFSKQOYNNIDV	1339
1321	SPARLAYQDKGVLHNEVKYSILWRGLPNVVTSAISLPNIRKPGDGY	YVYAFSKQOYNNIDV	1380
1340	PSSTARAITTRSGOTLSKWYNC	P	1363
1381	PSSTARAITTRSGOTLSKWYNC	P	1404

RESULT 3

RESOLUTION  
AAB60568

AAB60368  
IN AAB60568 standard: protein: 1404 AA.

ID	NAME	AGE	SEX	REL	STATUS	DATE	TIME	LOCATION	REMARKS
1	John Doe	35	M	Spouse	Married	2023-10-27	14:30	Room 101	Initial check-in
2	Jane Doe	32	F	Spouse	Married	2023-10-27	14:30	Room 101	Initial check-in
3	Emily White	28	F	Spouse	Married	2023-10-27	14:30	Room 102	Initial check-in
4	Michael Brown	40	M	Spouse	Married	2023-10-27	14:30	Room 102	Initial check-in
5	Sarah Green	30	F	Spouse	Married	2023-10-27	14:30	Room 103	Initial check-in
6	David Black	38	M	Spouse	Married	2023-10-27	14:30	Room 103	Initial check-in
7	Olivia Grey	25	F	Spouse	Married	2023-10-27	14:30	Room 104	Initial check-in
8	James Blue	45	M	Spouse	Married	2023-10-27	14:30	Room 104	Initial check-in
9	Ava Pink	22	F	Spouse	Married	2023-10-27	14:30	Room 105	Initial check-in
10	Noah Yellow	33	M	Spouse	Married	2023-10-27	14:30	Room 105	Initial check-in

XX  
X  
APR 05 68.

AC  
AB  
yy

XX  
DE 27-APR-2001 (first entry)

27-yy

XX  
DE  
Human monocyte-stimulating factor (MSF CACP)DE Hum  
vv[illegible]

Human; CACP protein; camptodactyly-arthropathy-coxa vara-p

KW MSF; megakaryocyte stimulating factor; synovial lubricant;

KW chromosome 1q2

KW ant

XX

SO  
HOM

XX

PN	WO2
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

XX

PD 01-

XX

PF 21-




PR 23-JUL-1999; 99US-0145328P.

PR 19-

XXXX

PA (UY)

XX  
F7  
10)

PT Warman ML:

XX  
TJ  
TSM

WPT: 2001-182721/18.

UR  
XX  
WET

XX  
PT  
New composition comprising the camptodactylv-arthropathv-coxa vara-

New composition comprising the camphorally-arachidonic ester, pericarditis protein in combination with an anesthetic useful for

pericarditis protein in combination with an anesthetic, associated with lubricants of tissue and joints.

xx  
xx

XX  
DD  
Brennan] c 1. Page: 34nn. English

Exa  
PS  
vv

The invention relates to a method of treating osteoarthritis via the administration of a composition comprising the camptodactyly-arthropathy-coxa vara-pericarditis (CACP) protein, or portions of the CACP protein. The composition may further comprise a local anaesthetic. The composition of the invention may be administered via intra-articular or intravenous injection. The human CACP protein is identified in the invention as being megakaryocyte stimulating factor (MSF). The gene encoding CACP protein (MSF) is located on chromosome 1q25-31, and mutations in this gene are responsible for the heritable disorder camptodactyly-arthropathy-coxa vara-pericarditis, in which patients have synovial hyperplasia without evidence of inflammation. CACP protein (MSF) acts as a synovium lubricant, and can be used to lubricate tissue and joints in the treatment of osteoarthritis. The composition may be applied to reduce the symptoms of osteoarthritis (e.g., joint pain, loss of range of movement or joint damage). The present sequence represents human megakaryocyte stimulating factor (MSF, CACP protein). Note: This sequence is not given in its entirety in figure 4 of the specification, although a GenBank accession number was given. This sequence was therefore obtained from GenBank (U70316)



QY 980 POKETKAPKPTSTKKPKTMRVRKPKTTPRKMSTMPBLNPTSRIAEAMLOTTTRPN 1039  
 DB 1021 POKETKAPKPTSTKKPKTMRVRKPKTTPRKMSTMPBLNPTSRIAEAMLOTTTRPN 1080  
 QY 1040 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMOYLPRVNPQGIINPMLS 1099  
 DB 1081 QTPNSKLVEVNPKESEDAGAGETPHMLLRPHVFMPEVTPDMOYLPRVNPQGIINPMLS 1140  
 QY 1100 DETNICNGKPVGDLTLRLNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSDIDTVFT 1159  
 DB 1141 DETNICNGKPVGDLTLRLNGTLVAFRGHYFWMLSPFSPSPARRITEVWGIPSDIDTVFT 1200  
 QY 1160 RCNCEGTFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGIIVAAALSTAKYKNWPSVY 1219  
 DB 1201 RCNCEGTFKDSQYWRFTNDIKDAGYKPIFKGFGGLTQGIIVAAALSTAKYKNWPSVY 1260  
 QY 1220 FEKRGGSIQYIKQEPVQKCPGRPALNYPVYGMTQVRRRFEAIGPSQTHIRIQY 1279  
 DB 1261 FEKRGGSIQYIKQEPVQKCPGRPALNYPVYGMTQVRRRFEAIGPSQTHIRIQY 1320  
 QY 1280 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPKNRPGDYVYAFSKDQYINIDV 1339  
 DB 1321 SPARLAYQDKGVLHNEVKVSIILWRGLPNVVTSAISLNPKNRPGDYVYAFSKDQYINIDV 1380  
 QY 1340 PSARAIAITRSGQTLSKWNVNC 1363  
 DB 1381 PSARAIAITRSGQTLSKWNVNC 1404

RESULT 4

ADM98014  
 ID ADM98014 standard; protein; 1404 AA.  
 AC ADM98014;  
 XX  
 DT 01-JUL-2004 (first entry)  
 XX Human megakaryocyte stimulating factor (MSF).  
 DE lubricating polypeptide; O-linked oligosaccharide; joint lubrication;  
 KW CAP; camptodactyl-arthropathy pericarditis; osteoarthritis; human;  
 KW megakaryocyte stimulating factor; MSF.  
 XX  
 OS Homo sapiens.  
 XX  
 XX US2004072741-A1.  
 XX  
 PD 15-APR-2004.  
 XX  
 PF 02-JUL-2001; 2001US-00897188.  
 XX  
 PR 23-APR-1999; 99US-00298970.  
 PR 24-APR-2000; 2000US-00556246.  
 XX  
 PA (JAYG/) JAY G D.  
 XX  
 PI Jay GD;  
 XX  
 DR WPI; 2004-373948/35.  
 DR N-PSDB; ADM98015.  
 XX

XX New tribonectin polypeptides and polynucleotides for lubricating joints  
 PT or other tissues to prevent or treat Camptodactyl-arthropathy-  
 PT pericarditis syndrome or osteoarthritis.  
 XX  
 PS Claim 1; SEQ ID NO 1; 34pp; English.  
 XX  
 CC The invention relates to a lubricating polypeptide and at least one O-  
 CC linked oligosaccharide. The composition and methods are useful for  
 CC lubricating joints or other tissues to prevent or treat camptodactyl-  
 CC arthropathy pericarditis (CAP) or osteoarthritis in mammals. The present  
 CC sequence represents the amino acid sequence of the human megakaryocyte

CC stimulating factor (MSF).  
 XX  
 SQ Sequence 1404 AA;  
 Query Match 99.8%; Score 7261.9; DB 8; Length 1404;  
 Best Local Similarity 97.1%; Pred. No. 7.3e-177;  
 Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;  
 QY 1 MAWKTLPIYLLLSVFVIQVSSQ----- 25  
 DB 1 MAWKTLPIYLLLSVFVIQVSSQ----- 60  
 QY 26 -----FLSCKGRCFESFERGREGCDCAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 79  
 DB 61 KRVCATLSCKGRCFESFERGREGCDCAQCKYDKCCPDYSEFCAEVHNPTSPSSKKAP 120  
 QY 80 PPSGASQTIKSTTKRSPKPNKKTKKVISEETIEHSVSENQESSSSSSSSSSSTI 139  
 DB 121 PPSGASQTIKSTTKRSPKPNKKTKKVISEETIEHSVSENQESSSSSSSSSTI 180  
 QY 140 KIKSSKNSAANRELQKKLVKDKNKKRKKKTPKPPVVDDEAGSLDNGDFKVTTPDTST 199  
 DB 181 KIKSSKNSAANRELQKKLVKDKNKKRKKKTPKPPVVDDEAGSLDNGDFKVTTPDTST 240  
 QY 200 TOHNVKYSTSPKITTAKPINPRPSLPNSDTSKETSITLVNKETTIVTKETTTNKQSTDG 259  
 DB 241 TOHNVKYSTSPKITTAKPINPRPSLPNSDTSKETSITLVNKETTIVTKETTTNKQSTDG 300  
 QY 260 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTPTTKEPAS 319  
 DB 301 KEKTTSAKETQSIEKTSADLAPTSKVLAKPTPKAETTTKGPALTTTKEPTPTTKEPAS 360  
 QY 320 TTPKEPTPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 379  
 DB 361 TTPKEPTPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 420  
 QY 380 APTTTSAPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 439  
 DB 421 APTTTSAPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 480  
 QY 440 EPAPTAPKPAPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 499  
 DB 481 EPAPTAPKPAPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 540  
 QY 500 TTKSAPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 559  
 DB 541 TTKSAPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 600  
 QY 560 APTAPKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTT 619  
 DB 601 APTAPKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTT 660  
 QY 620 PEEPAPTTPKAAAPNTPEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 679  
 DB 661 PEEPAPTTPKAAAPNTPEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPAS 720  
 QY 680 APTTPKAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKBPAPTTKEPASPTTKEPASPTT 739  
 DB 721 APTTPKAPKELAPTTTKEPTSTTSKDPAPTTPKGTAPTTPKBPAPTTKEPASPTTKEPASPTT 780  
 QY 740 TAPTTLKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTT 799  
 DB 781 TAPTTLKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTTKEPASPTT 840  
 QY 800 KPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPPDESTPELSAETPKALENSPKPGVPT 859  
 DB 841 KPAPTTPETPPPTTSEVSTPTTKEPTTIHKSPPDESTPELSAETPKALENSPKPGVPT 900  
 QY 860 TKTAAATKPEMTTAAKOKTTERDLRTTPTTAAAPKMTKETAATTEKTESKIATTTQV 919  
 DB 901 TKTAAATKPEMTTAAKOKTTERDLRTTPTTAAAPKMTKETAATTEKTESKIATTTQV 960  
 QY 920 TSTTTQDTPPKIITLTKTTLLAPKVTTTKTITTTTEIMNKPEETAAPKDRATNSKATTPK 979









QY 181 AGSGLDNGDFKVTTPDTSTTOHKNVSTSPKITTAKETINPRPSLPNSDTSKETSITVNKE 240  
DB 88 AGSGLDNGDFKVTTPDTSTTOHKNVSTSPKITTAKETINPRPSLPNSDTSKETSITVNKE 147  
QY 241 TTVETKETTITNKQSTDCGKETTSAKETQSTKETSADKOLAPTSKVLAKPTPKAETTTKG 300  
DB 148 TTVETKETTITNKQSTDCGKETTSAKETQSTKETSADKOLAPTSKVLAKPTPKAETTTKG 207  
QY 301 PALTTPEKPTPTTPKEPASPTTPKEPTTTIKSAPTTPKGPAPTTKSAPTTPKGPAPTTT 360  
DB 208 PALTTPEKPTPTTPKEPASPTTPKEPTTTIKSAPTTPKGPAPTTKSAPTTPKGPAPTTT 267  
QY 361 KEPAPTTPKEPAPTTTKSAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 420  
DB 268 KEPAPTTPKEPAPTTTKSAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 327  
QY 421 TTPKEPAPTTPKEPAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKE 480  
DB 328 TTPKEPAPTTPKEPAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKE 387  
QY 481 PAPTTTKSAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 540  
DB 388 PAPTTTKSAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 447  
QY 541 KEPAPTTPKEPAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 600  
DB 448 KEPAPTTPKEPAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 507  
QY 601 PTPPEELAPTTPEEPTPTTPKEEPAPTTPKAAAPNTPKGPAPTTTPKGPAPTTTPKGPAPTT 660  
DB 508 PTPPEELAPTTPEEPTPTTPKEEPAPTTPKAAAPNTPKGPAPTTTPKGPAPTTTPKGPAPTT 567  
QY 661 KETAPPTPKGAPTTLKEPAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 720  
DB 568 KETAPPTPKGAPTTLKEPAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 627  
QY 721 PKGPAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 780  
DB 628 PKGPAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 687  
QY 781 TPKEPTAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 840  
DB 688 TPKEPTAPTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTTTPKGPAPTT 747  
QY 841 AEPTPKALENSPKGPVPTTKPAATPKPMTTAKOKTTERDLRTTPTTAAAPKMTKET 900  
DB 748 AEPTPKALENSPKGPVPTTKPAATPKPMTTAKOKTTERDLRTTPTTAAAPKMTKET 807  
QY 901 ATTTKETSKITATTTQVSTTTQDTPPKITLTKTTLAPKVTITTKITITTEIMNKP 960  
DB 808 ATTTKETSKITATTTQVSTTTQDTPPKITLTKTTLAPKVTITTKITITTEIMNKP 867  
QY 961 EETAAPKDRATNSKATTPKQKPTKAPKPTSTKPKPTMPVRKPTTTPKMTSTWPE 1020  
DB 868 EETAAPKDRATNSKATTPKQKPTKAPKPTSTKPKPTMPVRKPTTTPKMTSTWPE 927  
QY 1021 LNPTSRJAEAMLQTTTRPNQTPNSKLVNPKSDAGAGETPHMLLRPHVFPVETPD 1080  
DB 928 LNPTSRJAEAMLQTTTRPNQTPNSKLVNPKSDAGAGETPHMLLRPHVFPVETPD 987  
QY 1081 MDYLPRVNOGIIINPMLSDETNI CNKFPVDGLTTLRNGTLVAFRGHYFWMLSFSPSP 1140  
DB 988 MDYLPRVNOGIIINPMLSDETNI CNKFPVDGLTTLRNGTLVAFRGHYFWMLSFSPSP 1047  
QY 1141 ARRITEVWGIPSPIDTVTRCNCEGKTFPFKDSQVWRFNTDIKDAGYKPKTFKFGGLGTG 1200  
DB 1048 ARRITEVWGIPSPIDTVTRCNCEGKTFPFKDSQVWRFNTDIKDAGYKPKTFKFGGLGTG 1107  
QY 1201 QIVAAALSTAKVNPESVYFPKRGSIQOYIYKQEPVQKCGRRPALNVPYVGETQVRR 1260  
DB 1108 QIVAAALSTAKVNPESVYFPKRGSIQOYIYKQEPVQKCGRRPALNVPYVGETQVRR 1167  
QY 1261 RFRFAIGPSQTHTIRIQSPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPIR 1320

DB 1168 RFRFAIGPSQTHTIRIQSPARLAYQDKGVHLNEVKVSIILWRGLPNVVTSAISLPIR 1227  
QY 1321 PDGYDYAFSKQYNNIDVPSRTARAITTRSGQTLISKVWYNCP 1363  
DB 1228 PDGYDYAFSKQYNNIDVPSRTARAITTRSGQTLISKVWYNCP 1270  
RESULT 9  
ADK67912  
ID ADK67912 standard; protein; 1311 AA.  
XX  
AC ADK67912;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Human extracellular messenger (EXMES) polypeptide.  
XX  
KW Human; extracellular messenger; EXMES; respiratory-gen.; anti-allergic;  
KW antiasthmatic; anti-inflammatory; antidiabetic; neuroprotective;  
KW muscular-gen.; antiarthritic; osteopathic; hepatotropic; antipsoriatic;  
KW varicide; fungicide; antiparasitic; protozoacide; antihelminthic;  
KW cytostatic; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PH Key Location/Qualifiers  
FT Peptide 1..30 /label= Signal peptide  
FT /note= "Spans residues 1 to 18, 20, 21, 24, 29 or 30 according to identification method"  
XX  
PN WO2004013292-A2.  
XX  
PD 12-FEB-2004.  
XX  
PF 30-JUL-2003; 2003WO-US024084.  
XX  
PR 02-AUG-2002; 2002US-0400810P.  
PR 19-SEP-2002; 2002US-0412197P.  
PR 04-OCT-2002; 2002US-0416004P.  
PR 08-NOV-2002; 2002US-0424862P.  
XX  
PA (INCY-) INCYTE CORP.  
XX  
PI Elliott VS, Khare R, Tran UK, Swarnakar A, Marquis JP;  
PI Richardson TW, Emerling BM, Lindquist EA, Chawla NK, Ramkumar J;  
PI Lee SY;  
XX  
WPI: 2004-157116/15.  
N-PSDB; ADK67917.  
XX  
New extracellular messengers and nucleic acids, useful for diagnosing, treating or preventing e.g. allergies, anemia, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoporosis, psoriasis, cancers or autoimmune thyroiditis.  
XX  
PS Claim 60; SEQ ID NO 5; 165pp; English.  
XX  
The present sequence is that of novel human extracellular messenger (EXMES) incyte ID NO: 7513018CD1 polypeptide. The protein shows homology to human megakaryocyte stimulating factor. The invention provides EXMES polynucleotides and polypeptides, as well as expression vectors, host cells, antibodies, agonists and antagonists, and methods for diagnosing, treating or preventing disorders associated with aberrant expression of EXMES, especially autoimmune and inflammatory disorders, cell proliferative disorders and endocrine disorders, e.g. adult respiratory distress syndrome, allergy, anaemia, erythroblastosis foetalis, Crohn's disease, diabetes mellitus, myasthenia gravis, osteoarthritis, osteoporosis, pancreatitis, psoriasis, viral, bacterial, fungal, parasitic, protozoal or helminthic infections, cancers, autoimmune thyroiditis, cretinism, Plummer's disease or thyroid carcinoma. Embodiments also provide methods for using the purified EXMES and/or

CC their encoding polynucleotides for facilitating the drug discovery  
 CC process, including determining of efficacy, dosage, toxicity and  
 CC pharmacology, and for investigating the pathogenesis of diseases and  
 CC medical conditions.  
 XX  
 XX  
 SQ Sequence 1311 AA;

Query Match 93.1%; Score 6774.6; DB 8; Length 1311;  
 Best Local Similarity 90.4%; Pred. No. 1.8e-164;  
 Matches 1269; Conservative 0; Mismatches 1; Indels 134; Gaps 2;

QY 1 MAWKTLPIYLILLLSVFVIOVSSQ----- 25  
 DB 1 MAWKTLPIYLILLLSVFVIOVSSQDLSSCAGRCGEGYRDATCNCVNCQHYMECCPDF 60  
 QY 26 -----ELSCGRGCFEFERGEDCDAQCKKYDKCCPDYSEFCAEYHNPTSPSSKKAP 79  
 DB 61 KRVCIAELSCGRGCFEFERGEDCDAQCKKYDKCCPDYSEFCAE----- 106  
 QY 80 PPSGASQTIKSTTKRSPKPNKXKTKVIESEBITEHSVSENQESSSSSSSSSSSTIW 139  
 DB 107 ----- 106  
 QY 140 KIKSSKNSAANRELQKKLVKONKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 199  
 DB 107 -----VKONKNRTKKKPTPKPPVDEAGSLDNGDFKVTTPDTST 147  
 QY 200 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTNNKQSTGDG 259  
 DB 148 TOHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETTVEKETTNNKQSTGDG 207  
 QY 260 KEKTSKAKTOSIEKTSKOLAPTSKVLAKPTKPAETTTKGPALTTKKEPTTTPKEPAS 319  
 DB 208 KEKTSKAKTOSIEKTSKOLAPTSKVLAKPTKPAETTTKGPALTTKKEPTTTPKEPAS 267  
 QY 320 TTPKEPTPTTIKSAPTTKPEPATTTKSAPTTKPEPATTTKGPATTTPKEPATTTKPE 379  
 DB 268 TTPKEPTPTTIKSAPTTKPEPATTTKSAPTTKPEPATTTKGPATTTPKEPATTTKPE 327  
 QY 380 APITTKSAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKSAPTTTKPEPATT 499  
 DB 388 EPAPTAPKKPAPTTPKEPATTTKPEPATTTKPEPATTTKPEPATTTKSAPTTTKPE 447  
 QY 500 TTKSAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 559  
 DB 448 TTKSAPTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTKPEPATTTK 507  
 QY 560 APTAPKEPATTTPKETAPTTPKLTPTTPEKLAPTTPEKAPATTPEELAPTTPEEPTPT 619  
 DB 508 APTAPKEPATTTPKETAPTTPKLTPTTPEKLAPTTPEKAPATTPEELAPTTPEEPTPT 567  
 QY 620 PREPAPTTPKAAAPNTKGPAPTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEP 679  
 DB 568 PREPAPTTPKAAAPNTKGPAPTTPKEPATTTPKEPATTTPKETAPTTPKGTAPTTLKEP 627  
 QY 680 APTTPKKKAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTPKEPATTTPKGPAPTTPKG 739  
 DB 628 APTTPKKKAPKELAPTTTKEPTSTSDKAPATTTPKGTAPTTPKEPATTTPKGPAPTTPKG 687  
 QY 740 TAPPTLKEPAPTTPKKPAPKELAPTTTKEPTSTSDKAPATTTPKETAPTTPKEPATTTPK 799  
 DB 688 TAPPTLKEPAPTTPKKPAPKELAPTTTKEPTSTSDKAPATTTPKETAPTTPKEPATTTPK 747  
 QY 800 KPAPTTPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPKALENPKBGPVPT 859  
 DB 748 KPAPTTPPTTSEVSTPTTTKEPTTIHKSPDESTPELSAEPKALENPKBGPVPT 807  
 QY 860 TKTPAAKPEMTTAKDTERDLRTTETTTAAPKMTKETATTTTEKTTESKITATTQV 919  
 DB 868 TKTPAAKPEMTTAKDTERDLRTTETTTAAPKMTKETATTTTEKTTESKITATTQV 867  
 QY 920 TSTTTQDTPPKITTLTKTTILAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATPK 979  
 DB 868 TSTTTQDTPPKITTLTKTTILAPKVTTTKKTIITTEIMNKPEETAKPKDRATNSKATPK 927  
 QY 980 POKPTKAPKPTSTKKPKTMRVRKPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 1039  
 DB 928 POKPTKAPKPTSTKKPKTMRVRKPKTTPTRKMTSTMPELNPTSRIAEAMLOTTTRPN 987  
 QY 1040 QTNSKLVNPNKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1099  
 DB 988 QTNSKLVNPNKSEDAGAGETPHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1047  
 QY 1100 DETNICNGKPVGDLTTLRNGTLVAFRGHYFWMLSFPSPSPARRITEVWGIPSPIDIVFT 1159  
 DB 1048 DETNICNGKPVGDLTTLRNGTLVAFRGHYFWMLSFPSPSPARRITEVWGIPSPIDIVFT 1107  
 QY 1160 RCNCEGKTFEFKSOYWRFTNDIKDAGYKPKPIFKGFGGLTGQIIVAALSTAKYKNWPSVY 1219  
 DB 1108 RCNCEGKTFEFKSOYWRFTNDIKDAGYKPKPIFKGFGGLTGQIIVAALSTAKYKNWPSVY 1167  
 QY 1220 FFKRGSSIOQYIYKQEPVQKCPGRPALNYPVVGEMTOVRRRRPERRAIGPSQTHIRIOY 1279  
 DB 1168 FFKRGSSIOQYIYKQEPVQKCPGRPALNYPVVGEMTOVRRRRPERRAIGPSQTHIRIOY 1227  
 QY 1280 SPARLAYQDGVHLNHNKVSILMRGLPNVVTSAISLNRKPKDGYDYAFSKDQYNNIDV 1339  
 DB 1228 SPARLAYQDGVHLNHNKVSILMRGLPNVVTSAISLNRKPKDGYDYAFSKDQYNNIDV 1287  
 QY 1340 PSRTARAITTRSGQTLISKVWYNCP 1363  
 DB 1288 PSRTARAITTRSGQTLISKVWYNCP 1311

RESULT 10  
 AAM24322  
 ID AAM24322 standard; protein; 1299 AA.  
 XX AAM24322;  
 AC  
 DT 12-OCT-2001 (first entry)  
 XX  
 DE Human EST encoded protein SEQ ID NO: 1847.  
 XX  
 KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KW diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;  
 KW gene therapy; nutrition.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200154477-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 25-JAN-2001; 2001WO-US002687.  
 XX  
 PR 25-JAN-2000; 2000US-00491404.  
 PR 17-JUL-2000; 2000US-00617746.  
 PR 03-AUG-2000; 2000US-00631451.  
 PR 15-SEP-2000; 2000US-00663870.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
 PI Cao Y, Drmanac RA, Zhang J, Wehrman T;  
 XX  
 DR WPI; 2001-476164/51.  
 DR N-PSDB; AAF98981.  
 XX  
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising  
 PT antibodies and research use.



XX PS Claim 20; Page 1198-1201; 1275pp; English.

XX CC The present invention provides the protein and coding sequences of novel

XX CC proteins from a variety of organisms, including human, dog, cat, horse,

XX CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

XX CC urchin and tomato. These were derived from expressed sequence tags (ESTs)

XX CC from the organism of interest. They can be used in diagnostics,

XX CC forensics, gene mapping, identification of mutations, to assess

XX CC biodiversity and for nutritional purposes. The present sequence is a

XX CC protein of the invention

XX SQ Sequence 1299 AA;

Query Match 91.9%; Score 6685.9; DB 4; Length 1299;

Best Local Similarity 96.8%; Pred. No. 3.3e-162;

Matches 1254; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MAWKTLPIVILLLLSVFVIQQVSSQ----- 25

DB 1 MAWKTLPIVILLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNCYDNCQHYMECCPDF 60

QY 26 -----ELSCGRCEFSFREGRECDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 79

DB 61 KRVTALSLCKGRCEFSFREGRECDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120

QY 80 PPSGASQTKSTTKRSPKPPNKKTKKVIIESEITEHSVSSENQESSSSSSSSSSSTIW 139

DB 121 PPSGASQTKSTTKRSPKPPNKKTKKVIIESEITEHSVSSENQESSSSSSSSSSSTIW 180

QY 140 KIKSSKNSAANRELQKLVKDNKKNTKKPKPPVDEAGSLDNGDFKVTTPDTST 199

DB 181 KIKSSKNSAANRELQKLVKDNKKNTKKPKPPVDEAGSLDNGDFKVTTPDTST 240

QY 200 TQNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKEITVETKETTINKQSTDG 259

DB 241 TQNKVSTSPKITTAKPINRPSLPNPSDTSKETSIVNKEITVETKETTINKQSTDG 300

QY 260 KEKTSAKETQSIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTTKEPTPTTPKEPAS 319

DB 301 KEKTSAKETQSIKTSKOLAPTSKVLAKPTPKAETTTKGPALTTTKEPTPTTPKEPAS 360

QY 320 TTPKEPTPTTIKSAPTTKPEAPTTTKSAPTTKPEAPTTTKEAPTTTKEAPTTTKEP 379

DB 361 TTPKEPTPTTIKSAPTTTKEAPTTTKSAPTTKPEAPTTTKEAPTTTKEAPTTTKEP 420

QY 380 APTTTKSAPTTKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 439

DB 421 APTTTKSAPTTKPEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 480

QY 440 EPAPTAPKPAPTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 499

DB 481 EPAPTAPKPAPTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 540

QY 500 TTKSAPTTKPEPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 559

DB 541 TTKSAPTTKPEPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 600

QY 560 APTAPKPAPTTKEPTTTPKLTPTTPEKLAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 619

DB 601 APTAPKPAPTTKEPTTTPKLTPTTPEKLAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 660

QY 620 PEEPAPTTPKAAAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 679

DB 661 PEEPAPTTPKAAAPTTPKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 720

QY 680 APTTPKPAKELAPTTTKEPTSTSDKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 739

DB 721 APTTPKPAKELAPTTTKEPTSTSDKEAPTTTKEAPTTTKEAPTTTKEAPTTTKEP 780

QY 740 TAPTTLKEPAPTTPKPAKELAPTTTKEPTSTSDKEAPTTTKEAPTTTKEAPTTTKEP 799

DB 781 TAPTTLKEPAPTTPKPAKELAPTTTKEPTSTSDKEAPTTTKEAPTTTKEAPTTTKEP 840

QY 800 KPAPTTPPTTSEVSTPTTTTKEPTTIHKSPDSTPELSAEPPTKALENSPKFPGVPT 859

DB 841 KPAPTTPPTTSEVSTPTTTTKEPTTIHKSPDSTPELSAEPPTKALENSPKFPGVPT 900

QY 860 TKTPAATKPEMTTAAKDKTTERDLRTTPTTTAAAPKMTKETATTTTETTESKITATTQV 919

DB 901 TKTPAATKPEMTTAAKDKTTERDLRTTPTTTAAAPKMTKETATTTTETTESKITATTQV 960

QY 920 TSTTQDTPPKITTLTKTTLAPKVTITTKTITTTIEMNKPETAKPKORATNSKATTPK 979

DB 961 TSTTQDTPPKITTLTKTTLAPKVTITTKTITTTIEMNKPETAKPKORATNSKATTPK 1020

QY 980 POKPTKAPKPTSTKPKTMPVRKPTTPTPKMTSTMPELNPTSRIAEAMLTQTTTPN 1039

DB 1021 POKPTKAPKPTSTKPKTMPVRKPTTPTPKMTSTMPELNPTSRIAEAMLTQTTTPN 1080

QY 1040 QTPNSKLVEVNPXSADAGAGETPHMLLRPHVFMPEVTPDMYLRVNPQGGIINPMLS 1099

DB 1081 QTPNSKLVEVNPXSADAGAGETPHMLLRPHVFMPEVTPDMYLRVNPQGGIINPMLS 1140

QY 1100 DETNICNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPEFSPSPARRITEVWGIPIPIDIVFT 1159

DB 1141 DETNICNGKPVVDGLTTLRNGTLVAFRGHYFWMLSPEFSPSPARRITEVWGIPIPIDIVFT 1200

QY 1160 RCNCEGKTPFFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYKNWPESVY 1219

DB 1201 RCNCEGKTPFFKDSQYWRFTNDIKDAGYKPIFKFGGLTGQIVAAALSTAKYKNWPESVY 1260

QY 1220 FFKRGSSIQQYIYKQBPVQKCPGRRPALNYPVYGE 1254

DB 1261 FFKRGSSIQQYIYKQBPVQKCPGRRPALNYPVYGE 1295

RESULT 11

AAB29778

ID AAB29778 standard; protein; 902 AA.

XX AC AAB29778;

XX DT 28-FEB-2001 (first entry)

XX DE Human MSF-derived tribonectin.

XX KM Human tribonectin; MSF; megakaryocyte stimulating factor;

XX KW alternative splicing; joint boundary lubricant; O-linked oligosaccharide;

XX KW osteoarthritis; tribosupplementation; tissue adhesion inhibition;

XX KW friction coefficient reduction; gene therapy; antiarthritic; osteopathic.

XX OS Homo sapiens.

XX FN WO200064930-A2.

XX PD 02-NOV-2000.

XX PF 24-APR-2000; 2000WO-US010953.

XX PR 23-APR-1999; 99US-00298970.

XX PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.

XX PI Jay GD;

XX DR WPI; 2001-024673/03.

XX PT Novel tribonectin polypeptide useful as lubricant for treating

XX PT osteoarthritis, comprises O-linked lubricating moiety.

XX PS Disclosure; Fig 1; 47pp; English.

XX CC The invention relates to a human tribonectin which is a product of

XX CC alternative splicing of the human MSF (megakaryocyte stimulating factor)

XX CC gene. The tribonectin has at least one O-linked oligosaccharide





Query Match 40.3%; Score 2929; DB 4; Length 551;  
Best Local Similarity 100.0%; Pred. No. 6.8e-67; Indels 0; Gaps 0;  
Matches 551; Conservative 0; Mismatches 0;

QY 556 TKKPAPTAPKEPAPTPKETAAPTPKPLTPPTPEKLAAPTPBKPAAPTPPEELAPTPPEP 615  
DB 1 TKKPAPTAPKEPAPTPKETAAPTPKPLTPPTPEKLAAPTPBKPAAPTPPEELAPTPPEP 60

QY 616 TPTTPEEPAPTPKAAAPNTPKEPAPTPKPAAPTPKPAAPTPKETAAPTPKGTAPTT 675  
DB 61 TPTTPEEPAPTPKAAAPNTPKEPAPTPKPAAPTPKPAAPTPKETAAPTPKGTAPTT 120

QY 676 LKEPAPTPKPAKAPKELAPTTTKEPTSTSDKPAAPTPKGTAPTPKPAAPTPKPAAP 735  
DB 121 LKEPAPTPKPAKAPKELAPTTTKEPTSTSDKPAAPTPKGTAPTPKPAAPTPKPAAP 180

QY 736 TPKGTAPTTLKEPAPTPKPAKAPKELAPTTTKEPTSTSDKPAAPTPKETAAPTPKPAAP 795  
DB 181 TPKGTAPTTLKEPAPTPKPAKAPKELAPTTTKEPTSTSDKPAAPTPKETAAPTPKPAAP 240

QY 796 TTPKKAPAPTPPEPPTTSVSTPTTKEPTTIHKSPDESTPELSAEPKPALENSPKPEP 855  
DB 241 TTPKKAPAPTPPEPPTTSVSTPTTKEPTTIHKSPDESTPELSAEPKPALENSPKPEP 300

QY 856 GVPETTKTPAATKEPMITTAKDKTTERDLRTTPETTTAAAPKMTKETATTTTEKTTESKITAT 915  
DB 301 GVPETTKTPAATKEPMITTAKDKTTERDLRTTPETTTAAAPKMTKETATTTTEKTTESKITAT 360

QY 916 TTQVSTSTTQDTPPFKITTLTKTTLAPKVTTTKKTTITTEIMNKPEETAKPKDRATNSKA 975  
DB 361 TTQVSTSTTQDTPPFKITTLTKTTLAPKVTTTKKTTITTEIMNKPEETAKPKDRATNSKA 420

QY 976 TTPKPKQPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPBLNPTSRIAEAMLQTT 1035  
DB 421 TTPKPKQPTKAPKKPTSTKKPKTMPRVRKPKTTPTPRKMTSTMPBLNPTSRIAEAMLQTT 480

QY 1036 TRNQTPNSKLVENPKSEDAAGGEGTPEHMLLRPHVFMPEVTPDMDYLRVFNQGLIIN 1095  
DB 481 TRNQTPNSKLVENPKSEDAAGGEGTPEHMLLRPHVFMPEVTPDMDYLRVFNQGLIIN 540

QY 1096 PMLSDETINICN 1106  
DB 541 PMLSDETINICN 551

RESULT 13  
ABUS3252  
ID ABUS3252 standard; protein; 546 AA.  
XX AC ABUS3252;

XX DT 14-APR-2003 (first entry)  
XX DE Human testes-derived DKFZphtes3\_4o19 homologue #1.  
XX KW Human; gene therapy; vaccine; disease treatment; detection.  
XX OS Homo sapiens.

XX PN WO200112659-A2.  
XX PD 22-FEB-2001.

XX PF 18-AUG-2000; 2000WO-IB001496.  
XX PR 18-AUG-1999; 99US-0149499P.  
XX ER 28-SEP-1999; 99US-0156503P.  
XX XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX FA Wiemann S;  
XX PI  
XX XX WPI; 2001-327840/34.  
XX DR

XX Nucleic acids having the sequences of clones isolated from libraries of  
PT different human tissues, useful in recombinant DNA methodologies.  
XX Example III; Page 892; 1095pp; English.  
XX This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence represents a homologue  
CC of a polypeptide described in the disclosure of the invention  
XX Sequence 546 AA;

Query Match 40.1%; Score 2920; DB 4; Length 546;  
Best Local Similarity 100.0%; Pred. No. 1.1e-66;  
Matches 546; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 447 KKPAPTPKPAAPTPKPAAPTPKPSPTTPKPAAPTTTKSAPTTTKSAPTTTKSAPT 506  
DB 1 KKPAPTPKPAAPTPKPAAPTPKPSPTTPKPAAPTTTKSAPTTTKSAPTTTKSAPT 60

QY 507 TPKERSPTTKKPAAPTPKPAAPTPKPAAPTPKPAAPTPKPAAPTTTKKPAAPKE 566  
DB 61 TPKERSPTTKKPAAPTPKPAAPTPKPAAPTPKPAAPTPKPAAPTTTKKPAAPKE 120

QY 567 PAPTTPKETAPTPKPLTPTTPEKLAAPTPPEKAPTTPEELAPTTPEEPPTTPEEPAPT 626  
DB 121 PAPTTPKETAPTPKPLTPTTPEKLAAPTPPEKAPTTPEELAPTTPEEPPTTPEEPAPT 180

QY 627 TPXAAAPNTPKPAAPTPKPAAPTPKPAAPTPKPAAPTPKPAAPTTTKKPAAPKE 686  
DB 181 TPXAAAPNTPKPAAPTPKPAAPTPKPAAPTPKPAAPTTTKKPAAPTTTKK 240

QY 687 PAKELAPTTTKEPTSTTSKPAAPTPKGTAPTPKPAAPTPKPAAPTTTKGTAAPTTLK 746  
DB 241 PAKELAPTTTKEPTSTTSKPAAPTPKGTAPTPKPAAPTPKPAAPTTTKGTAAPTTLK 300

QY 747 EPAPTTTPKPAKELAPTTTKGPTSTTSKPAAPTPKETAAPTPKPAAPTPKPAAPTTTP 806  
DB 301 EPAPTTTPKPAKELAPTTTKGPTSTTSKPAAPTPKETAAPTPKPAAPTTTPKPAAPTTTP 360

QY 807 ETDPPTTSVSTPTTKEPTTIHKSPDESTPELSAEPKPALENSPKPEPGVPTTKTAPAT 866  
DB 361 ETDPPTTSVSTPTTKEPTTIHKSPDESTPELSAEPKPALENSPKPEPGVPTTKTAPAT 420

QY 867 KPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETATTTTEKTTESKITATTTQVSTTTQD 926  
DB 421 KPEMTTAKDKTTERDLRTTPETTTAAAPKMTKETATTTTEKTTESKITATTTQVSTTTQD 480

QY 927 TTPFKITTLKTTTLAPKVTTTKKTTITTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 986  
DB 481 TTPFKITTLKTTTLAPKVTTTKKTTITTEIMNKPEETAKPKDRATNSKATTPKPKPTKA 540

QY 987 PKKPTS 992  
DB 541 PKKPTS 546

RESULT 14  
AAO18834  
ID AAO18834 standard; protein; 538 AA.  
XX AC AAO18834;  
XX XX 29-OCT-2002 (first entry)  
XX DT  
XX DE 3' cartilage superficial zone protein coding sequence encoded protein.

XX SZP; superficial zone protein; cartilage; lubrication; human;  
 KW degenerative joint condition; arthritis; osteoporosis; trauma; CACP;  
 KW chondroitin sulphate substitution consensus; antiarthritis;  
 KW antirheumatic; osteopathic; antigout; antiinflammatory; dermatological;  
 KW immunosuppressive.  
 XX Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 PH Misc-difference 114  
 FT /note= "encoded by ACTACT"  
 XX  
 PN WO200262847-A2.  
 XX  
 XX 15-AUG-2002.  
 XX  
 XX 31-DEC-2001; 2001WO-US050379.  
 XX  
 XX 29-DEC-2000; 2000US-0258920P.  
 XX  
 XX (GLAXO) GLAXO GROUP LTD.  
 PA (RUSH-) RUSH PRESBYTERIAN ST LUKE MEDICAL CENT.  
 XX  
 XX Hutchins JT, Kuettner KE, Schmid TM, Schumacher BL, Su J;  
 PI Dixon EP;  
 XX  
 XX WPI; 2002-636585/68.  
 DR N-PSDB; AAL49079.  
 XX  
 XX New purified superficial zone protein (SZP) polypeptides, useful for  
 PT treating degenerative joint conditions, e.g. osteoarthritis, rheumatoid  
 PT arthritis, gout, spondylarthritis, synovitis, tendonitis, lupus, or  
 PT osteoporosis.  
 XX  
 XX Claim 59; Page 86-87; 89pp; English.  
 XX  
 XX The present invention provides the protein and coding sequences of human  
 CC superficial zone protein (SZP). The protein is involved in the  
 CC lubrication of joints, and the sequences can be used in the treatment of  
 CC degenerative joint conditions or to delay symptoms of a degenerative  
 CC joint condition, e.g. osteoarthritis, rheumatoid arthritis, gout,  
 CC psoriatic arthritis, reactive arthritis, viral or post viral arthritis,  
 CC spondylarthritis, juvenile arthritis, synovitis, tendonitis, systemic  
 CC lupus erythematosus, CACP, osteoporosis or trauma. The present sequence  
 CC is the protein encoded by the human 3' cartilage SZP cDNA  
 XX  
 XX Sequence 538 AA;  
 XX  
 XX Query Match 39.2%; Score 2850.9; DB 5; Length 538;  
 XX Best Local Similarity 99.8%; Pred. No. 6.5e-65;  
 XX Matches 538; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 XX  
 QY 825 PTTIHKSPDESPFELSAETPKALENSPKPEPGVPTTKTPAATKPEMTTAKDKTTERDLR 884  
 Db 1 PTTIHKSPDESPFELSAETPKALENSPKPEPGVPTTKTPAATKPEMTTAKDKTTERDLR 60  
 QY 885 TTPETTTAAPKMTKETATTTETTESKITATTQVTSSTTQDTPFKITLTKTLAPKV 944  
 Db 61 TTPETTTAAPKMTKETATTTETTESKITATTQVTSSTTQDTPFKITLTK-TTLAPKV 119  
 QY 945 TTTTKTITTTETIMNKEETAKPKDRATNSKATTPPKPKPTKAPKPTSTKPKPTWPRVK 1004  
 Db 120 TTTTKTITTTETIMNKEETAKPKDRATNSKATTPPKPKPTKAPKPTSTKPKPTWPRVK 179  
 QY 1005 PKTTTPRKMTSTMPELNPTSIAEAMLTQTTTRPNQTPNSKLVEVNPKSSEDAGGAEGETP 1064  
 Db 180 PKTTTPRKMTSTMPELNPTSIAEAMLTQTTTRPNQTPNSKLVEVNPKSSEDAGGAEGETP 239  
 QY 1065 HMLLRPHVMPVETPDMVDYLPVPVNOGIIINPMLSDETNI CNGKPVGDGLTTILNGLTVAF 1124  
 Db 240 HMLLRPHVMPVETPDMVDYLPVPVNOGIIINPMLSDETNICNGKPVGDGLTTILNGLTVAF 299

QY 1125 RGHYFWMLSPPSPARRITEVWGIPSPIDTFTVTRCNCEGKTFFFKDSQVWRFTNDIKD 1184  
 Db 300 RGHYFWMLSPPSPARRITEVWGIPSPIDTFTVTRCNCEGKTFFFKDSQVWRFTNDIKD 359  
 QY 1185 AGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYEFKRGSGIOQYIYKQSPVQKCPGRR 1244  
 Db 360 AGYKPIFKGFGGLTGQIVAAALSTAKYKNWPESVYEFKRGSGIOQYIYKQSPVQKCPGRR 419  
 QY 1245 PALNYPVYGMTQVRRRRFERAIGPSQTHIRIQYSPARLAYQDKGVHNEVKVSIILWRG 1304  
 Db 420 PALNYPVYGMTQVRRRRFERAIGPSQTHIRIQYSPARLAYQDKGVHNEVKVSIILWRG 479  
 QY 1305 LPNVVTSAISLPNIRKPDGDYDYAFSKDQYINIDVPSRTARAITTRSGQTLKVVYNCP 1363  
 Db 480 LPNVVTSAISLPNIRKPDGDYDYAFSKDQYINIDVPSRTARAITTRSGQTLKVVYNCP 538  
 XX  
 XX RESULT 15  
 XX ABUS3254  
 XX ID ABUS3254 standard; protein; 513 AA.  
 XX AC ABUS3254;  
 XX DT 14-APR-2003 (first entry)  
 XX DE Human testes-derived DKFZphtes3\_4o19 homologue #3.  
 XX KW Human; gene therapy; vaccine; disease treatment; detection.  
 XX OS Homo sapiens.  
 XX PN WO200112659-A2.  
 XX PD 22-FEB-2001.  
 XX PF 18-AUG-2000; 2000WO-IB001496.  
 XX PR 18-AUG-1999; 99US-0149499P.  
 XX 28-SEP-1999; 99US-0156503P.  
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 XX Wiemann S;  
 XX WPI; 2001-327840/34.  
 XX Nucleic acids having the sequences of clones isolated from libraries of  
 XX different human tissues, useful in recombinant DNA methodologies.  
 XX Example III; Page 893; 1095pp; English.  
 XX This invention describes novel polynucleotides and polypeptides isolated  
 XX from human cDNA libraries which can be used for gene therapy or in  
 XX vaccines. The polynucleotides of the invention and antibodies encoded by  
 XX them may be used in the prevention, diagnosis and treatment of diseases  
 XX associated with inappropriate polypeptide expression. The products of the  
 XX invention may also be used to identify modulators of expression and  
 XX activity and to down regulate expression and activity. The antibodies of  
 XX the invention may also be used as diagnostic agents for detecting the  
 XX presence of polypeptides in samples. This sequence represents a homologue  
 XX of a polypeptide described in the disclosure of the invention  
 XX  
 XX Sequence 513 AA;  
 XX  
 XX Query Match 37.9%; Score 2757; DB 4; Length 513;  
 XX Best Local Similarity 100.0%; Pred. No. 1.5e-62;  
 XX Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 166 RTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHNKSVSTPKITTAKPINRPSLPP 225  
 Db 1 RTKKKPTKPPVDEAGSLDNGDFKVTTPDTSTTQHNKSVSTPKITTAKPINRPSLPP 60  
 QY 226 NSDTSKETSILVNKETTETTTTNTKQTDGKEKTSKETSQTSIEKTSKADLAPTSK 285

Db	61	NSDTSKETSLTVNKETTVETKTTTTNNKOTSIDGKEKTTSAKETQSIENISAKDLAPT	120
Qy	286	VLAKEPTPKAETTTKGPALTTPKETPTTTPKEPASATTPKEPTTTIKSAPTTPKEPAPTTT	345
Db	121	VLAKEPTPKAETTTKGPALTTPKETPTTTPKEPASATTPKEPTTTIKSAPTTPKEPAPTTT	180
Qy	346	KGAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT	405
Db	181	KGAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT	240
Qy	406	TPKEPAPTTTPKEPTPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP	465
Db	241	TPKEPAPTTTPKEPTPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEP	300
Qy	466	APTTPKEPSPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK	525
Db	301	APTTPKEPSPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPK	360
Qy	526	EPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT	585
Db	361	EPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT	420
Qy	586	TTPEKLAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE	645
Db	421	TTPEKLAPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE	480
Qy	646	EPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT	705
Db	481	EPAPTTPKPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT	540

Search completed: October 13, 2004, 11:37:15  
Job time : 116.889 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 13, 2004, 11:23:49 ; Search time 23.693 Seconds  
(without alignments)  
3815.116 Million cell updates/sec

Title: SEQ1-F  
Perfect score: 7276  
Sequence: 1 MAWKTPYVLLLSVFIQ.....ARAITRSGQTLKWNVNC 1363

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/iaa/5A COMB.pep.\*  
2: /cgn2\_6/prodata/1/iaa/5B COMB.pep.\*  
3: /cgn2\_6/prodata/1/iaa/6A COMB.pep.\*  
4: /cgn2\_6/prodata/1/iaa/6B COMB.pep.\*  
5: /cgn2\_6/prodata/1/iaa/6C COMB.pep.\*  
6: /cgn2\_6/prodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7276	100.0	1363	4	US-07-757-022B-52
2	7261.9	99.8	1404	4	US-07-757-022B-2
3	7261.9	99.8	1404	4	US-07-757-022B-62
4	7261.9	99.8	1404	4	US-09-298-970A-1
5	7232.9	99.4	1404	4	US-10-164-595-78
6	7058.7	97.0	1320	4	US-07-757-022B-46
7	7058.7	97.0	1320	4	US-07-757-022B-60
8	7044.6	96.8	1361	4	US-07-757-022B-40
9	7043.7	96.8	1320	4	US-10-164-595-58
10	7002	96.2	1313	4	US-07-757-022B-142
11	6987.9	96.0	1354	4	US-07-757-022B-48
12	6878.9	94.5	1314	4	US-07-757-022B-50
13	6794.7	93.4	1270	4	US-07-757-022B-44
14	6780.6	93.2	1311	4	US-07-757-022B-42
15	5820.9	80.0	1140	4	US-07-757-022B-104
16	5561	76.4	1049	4	US-07-757-022B-58
17	5495.5	75.5	1038	4	US-07-757-022B-74
18	5231.9	71.9	1022	4	US-07-757-022B-84
19	5011	68.9	941	4	US-07-757-022B-34
20	2160.9	29.7	422	4	US-07-757-022B-68
21	2146.8	29.5	463	4	US-07-757-022B-54
22	2032.8	27.9	423	4	US-07-757-022B-66
23	1886.9	25.9	372	4	US-07-757-022B-64
24	1371	18.8	5179	4	US-09-538-092-1258
25	1200.1	16.5	296	4	US-07-757-022B-70
26	1196.3	16.4	8991	4	US-08-714-741-32
27	946.2	13.0	237	4	US-07-757-022B-72

RESULT 1  
US-07-757-022B-52  
; Sequence 52, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1363 amino acids  
; TYPE: AMINO ACID

## ALIGNMENTS

Sequence 96, Appl  
Sequence 2, Appl  
Sequence 1, Appl  
Sequence 98, Appl  
Sequence 22, Appl  
Sequence 21, Appl  
Sequence 76, Appl  
Sequence 116, App  
Sequence 136, App  
Sequence 94, Appl  
Sequence 30, Appl  
Sequence 132, App  
Sequence 1142, Ap  
Sequence 92, Appl  
Sequence 5, Appl  
Sequence 5, Appl  
Sequence 425, App  
Sequence 5, Appl

! TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-52

Query Match 100.0%; Score 7276; DB 4; Length 1363;  
Best Local Similarity 100.0%; Pred. No. 1.6e-196;  
Matches 1363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAWKTLPIYLLLSLVFVIQVSSQBELSCKGRFCFESFERGRCDCDQAQCKYDKCCPDYE	60
DB	1	MAWKTLPIYLLLSLVFVIQVSSQBELSCKGRFCFESFERGRCDCDQAQCKYDKCCPDYE	60
QY	61	SFCAEVHNTSPSSKAPPPGASQTIKSTTKSPKPNKKTKKVIKVESEIIEBHSVS	120
DB	61	SFCAEVHNTSPSSKAPPPGASQTIKSTTKSPKPNKKTKKVIKVESEIIEBHSVS	120
QY	121	ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDNKNRKKKPTPKPPVDE	180
DB	121	ENQESSSSSSSSSTIWKIKSSKNSAANRELQKLVKDNKNRKKKPTPKPPVDE	180
QY	181	AGSLDNGDFKVTPTDTSITQHNKSVSPKITTAKDINPRPSLPNDSKETSITVNKE	240
DB	181	AGSLDNGDFKVTPTDTSITQHNKSVSPKITTAKDINPRPSLPNDSKETSITVNKE	240
QY	241	TTVETKETTNNKQTSIDGKEKTSKETSIAKTSKOLAPTSKVLAKPTPKAETTKG	300
DB	241	TTVETKETTNNKQTSIDGKEKTSKETSIAKTSKOLAPTSKVLAKPTPKAETTKG	300
QY	301	PALTTPKEPTTPPKPEASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTKPEAPTTT	360
DB	301	PALTTPKEPTTPPKPEASTTPKEPTTIKSAPTTPKEPAPTTTKSAPTTKPEAPTTT	360
QY	361	KEPAPTTKPEAPTTTKPEAPTTKSAPTTPKEPAPTTPKKAPPTPKPEAPTTKPEPT	420
DB	361	KEPAPTTKPEAPTTTKPEAPTTKSAPTTPKEPAPTTPKKAPPTPKPEAPTTKPEPT	420
QY	421	TTPEKAPTTKPEAPTTKPEAPTAAPKAPPTPKPEAPTTKPEAPTTTKESPTTKE	480
DB	421	TTPEKAPTTKPEAPTTKPEAPTAAPKAPPTPKPEAPTTKPEAPTTTKESPTTKE	480
QY	481	PAPTTTKSAPTTTKPEAPTTKSAPTTPKEPSPPTTKPEAPTTKPEAPTTPKKAPPT	540
DB	481	PAPTTTKSAPTTTKPEAPTTKSAPTTPKEPSPPTTKPEAPTTKPEAPTTPKKAPPT	540
QY	541	KEPAPTTKPEAPTTTKKAPTAAPKPEAPTTPKETAPTTPKLTPTTPEKLAAPTTPKPA	600
DB	541	KEPAPTTKPEAPTTTKKAPTAAPKPEAPTTPKETAPTTPKLTPTTPEKLAAPTTPKPA	600
QY	601	PTTPEELAPTTPEPTTPEEAPTTKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTT	660
DB	601	PTTPEELAPTTPEPTTPEEAPTTKAAAPNTKPEAPTTKPEAPTTKPEAPTTKPEAPTT	660
QY	661	KETAPTTKGTAPTTKPEAPTTKPEAPKELAPTTTKETSTTSKPEAPTTKGTAPT	720
DB	661	KETAPTTKGTAPTTKPEAPTTKPEAPKELAPTTTKETSTTSKPEAPTTKGTAPT	720
QY	721	KPEAPTTKPEAPTTKGTAPTTKPEAPTTKPEAPKELAPTTTKGTSTTSKPEAPT	780
DB	721	KPEAPTTKPEAPTTKGTAPTTKPEAPTTKPEAPKELAPTTTKGTSTTSKPEAPT	780
QY	781	TPKETAPTTPKEAPTTKPEAPTTBTPPTTSEVSTPTTKETPTTIHKSPDESTPELS	840
DB	781	TPKETAPTTPKEAPTTKPEAPTTBTPPTTSEVSTPTTKETPTTIHKSPDESTPELS	840
QY	841	AEPTPKALENSKPEPGVPTTKTAAKPEMTTTAKDKITERDLRTTPTTTAAPKMTKET	900
DB	841	AEPTPKALENSKPEPGVPTTKTAAKPEMTTTAKDKITERDLRTTPTTTAAPKMTKET	900
QY	901	ATTTEKTESKITATTQVSTTQDTPPKITLKITLAPKVTITTKITTTIEMNKP	960
DB	901	ATTTEKTESKITATTQVSTTQDTPPKITLKITLAPKVTITTKITTTIEMNKP	960
QY	961	BETAKPKDRATNSKATTTPKQKPTKAPKKPTSTKKPTMPRVKPTTTPPKMTSTMP	1020

DB	961	BETAKPKDRATNSKATTTPKQKPTKAPKKPTSTKKPTMPRVKPTTTPPKMTSTMP	1020
QY	1021	LNPTSRIAEAMLOTTTRNQTPNSKLVEVNPKESEDAGGAEGETHMLLRPHVFMPEVTPD	1080
DB	1021	LNPTSRIAEAMLOTTTRNQTPNSKLVEVNPKESEDAGGAEGETHMLLRPHVFMPEVTPD	1080
QY	1081	MDYLPRVNOGIIINPMLSDETNIENGKVDGLTTLRNCTLVAFRGHYFWMLSPFSPSP	1140
DB	1081	MDYLPRVNOGIIINPMLSDETNIENGKVDGLTTLRNCTLVAFRGHYFWMLSPFSPSP	1140
QY	1141	ARRITEVWGIPSPIDTFTFRCNCEGKTFEKKDQSVWRFNIDKAGYKPIFKFGGLTG	1200
DB	1141	ARRITEVWGIPSPIDTFTFRCNCEGKTFEKKDQSVWRFNIDKAGYKPIFKFGGLTG	1200
QY	1201	QIVAAALSTAKYKNWESVYFFKRGSSIQQYIYKQEPVQKCGRRPALNYPVYGEVQVRR	1260
DB	1201	QIVAAALSTAKYKNWESVYFFKRGSSIQQYIYKQEPVQKCGRRPALNYPVYGEVQVRR	1260
QY	1261	RRFERAIGPSQTHIRIQYSPARLAYQDKGVHLHNEVKVLSILWRGLPNVVTSAISLPIRK	1320
DB	1261	RRFERAIGPSQTHIRIQYSPARLAYQDKGVHLHNEVKVLSILWRGLPNVVTSAISLPIRK	1320
QY	1321	PDGYDYAFSKQYQYINIDVPSRTARAITTRSGQTLISKWYNCP	1363
DB	1321	PDGYDYAFSKQYQYINIDVPSRTARAITTRSGQTLISKWYNCP	1363

RESULT 2  
US-07-757-022B-2  
; Sequence 2, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luan  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170

TELEFAX: (617) 876-5851  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1404 amino acids  
 ; TYPE: AMINO ACID  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-07-757-022B-2

Query Match 99.8%; Score 7261.9; DB 4; Length 1404;  
 Best Local Similarity 97.18; Pred. No. 4.2e-196;  
 Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;

QY 1 MANKTLPIYLLLSLVFVIQVSSQ----- 25  
 DB 1 MANKTLPIYLLLSLVFVIQVSSQDLSSCAGRCGEGYSRDATCNDYCNQHYMECCPDF 60

QY 26 -----ELSCKGRCFESFERGECDDAQCCKYDKCCPDYESFCAEVHNFTSPSSKKAP 79  
 DB 61 KRVCATLSCKGRCFESFERGECDDAQCCKYDKCCPDYESFCAEVHNFTSPSSKKAP 120

QY 80 PPSGASQTIKSTTKRSPPKPKKKTKVIESEETEEHSYSENQESSSSSSSSSSSTIW 139  
 DB 121 PPSGASQTIKSTTKRSPPKPKKKTKVIESEETEEHSYSENQESSSSSSSSSSSTIW 180

QY 140 KIKSSKNSAANRELQKKLVKDKNKNRTKKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST 199  
 DB 181 KIKSSKNSAANRELQKKLVKDKNKNRTKKKPTKPPVVDVDEAGSLDNGDFKVTTPDTST 240

QY 200 TQHNKYSTSPKITTAKPINRPSLPNDSKTSKTSITVKNKETTVEKTTTNKQSTSDG 259  
 DB 241 TQHNKYSTSPKITTAKPINRPSLPNDSKTSKTSITVKNKETTVEKTTTNKQSTSDG 300

QY 260 KEKTTISAKETQSIEKTSADLAPTSKVLAKPTKAEITTKGPAITTPKEPTTPPKEPAS 319  
 DB 301 KEKTTISAKETQSIEKTSADLAPTSKVLAKPTKAEITTKGPAITTPKEPTTPPKEPAS 360

QY 320 TTPKEPTPTTKSAPTTPKEPATPTTKSAPTTPKEPATPTTKEPATPTTKEPATPTTKEP 379  
 DB 361 TTPKEPTPTTKSAPTTPKEPATPTTKSAPTTPKEPATPTTKEPATPTTKEPATPTTKEP 420

QY 380 APTTKSAPTTPKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTK 439  
 DB 421 APTTKSAPTTPKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTK 480

QY 440 EPAPTAPKKPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPAT 499  
 DB 481 EPAPTAPKKPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPAT 540

QY 500 TTKSAPTTPKEPSPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKKP 559  
 DB 541 TTKSAPTTPKEPSPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKEPATPTTKKP 600

QY 560 APTAPKEPATPTTKETAPTKLPTTPEKLAPTTPKEPATPTTKEPATPTTKETAPTKLPTT 619  
 DB 601 APTAPKEPATPTTKETAPTKLPTTPEKLAPTTPKEPATPTTKETAPTKLPTTPEKLAPTT 660

QY 620 PEEPATPTTKAAAPNTKEPATPTTKEPATPTTKEPATPTTKETAPTKLPTTPEKLAPTT 679  
 DB 661 PEEPATPTTKAAAPNTKEPATPTTKEPATPTTKEPATPTTKETAPTKLPTTPEKLAPTT 720

QY 680 APTTPKPKAPKELAPTTTKEPTSTISDKPATPTTKGTAPTTPKEPATPTTKEPATPTTKG 739  
 DB 721 APTTPKPKAPKELAPTTTKEPTSTISDKPATPTTKGTAPTTPKEPATPTTKEPATPTTKG 780

QY 740 TAPTTLKEPATPTTKAPKAPKELAPTTTGTSTSDKAPATPTTKETAPTKLPTTPEKLAPTT 799  
 DB 781 TAPTTLKEPATPTTKAPKAPKELAPTTTGTSTSDKAPATPTTKETAPTKLPTTPEKLAPTT 840

QY 800 KPAPTTPTPTPTTSEVSTPTTKETPTTIHKSPDESTPELSAAPTTPKALENSPKEPGVT 859  
 DB 841 KPAPTTPTPTPTTSEVSTPTTKETPTTIHKSPDESTPELSAAPTTPKALENSPKEPGVT 900

## RESULT 3

US-07-757-022B-62  
 ; Sequence 62, Application US/07757022B  
 ; Patent No. 6433142  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gesner, Thomas G.  
 ; APPLICANT: Clark, Stephen C.  
 ; APPLICANT: Turner, Katherine  
 ; APPLICANT: Hewick, Rodney M.  
 ; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
 ; NUMBER OF SEQUENCES: 143  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Genetics Institute, Inc.  
 ; STREET: 87 CambridgePark Drive  
 ; CITY: Cambridge  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02140  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/757,022B  
 ; FILING DATE: 19910910  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/643,502  
 ; FILING DATE: 18-JAN-1991  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/546,114  
 ; FILING DATE: 29-JUN-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/457,196  
 ; FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:		
; APPLICATION NUMBER: US 07/390,901		
; FILING DATE: 08-AUG-1989		
; ATTORNEY/AGENT INFORMATION:		
; NAME: Cseerr, Luann		
; REGISTRATION NUMBER: 31,822		
; REFERENCE/DOCKET NUMBER: GI 5190		
; TELECOMMUNICATION INFORMATION:		
; TELEPHONE: (617)876-1170		
; TELEFAX: (617)876-5851		
; INFORMATION FOR SEQ ID NO: 62:		
; SEQUENCE CHARACTERISTICS:		
; LENGTH: 1404 amino acids		
; TYPE: AMINO ACID		
; TOPOLOGY: linear		
; MOLECULE TYPE: protein		
US-07-757-022B-62		
Query Match 99.8%; Score 7261.9; DB 4; Length 1404;		
Best Local Similarity 97.1%; Pred. No. 4.2e-196;		
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;		
QY	1	MAWKTLPIYLILLLSVFVIQVSSQ-----25
Db	1	MAWKTLPIYLILLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY	26	-----ELSCCKGRCFESFERGREGDCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
Db	61	KRVCTAELSCGRCFESFERGREGDCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 120
QY	80	PPSGASQTKSTTKRSPKPNKKTKVIESEBITEHSHSVSENOESSSSSSSSSTIW 139
Db	121	PPSGASQTKISTTKRSPKPNKKTKVIESEBITEHSHSVSENOESSSSSSSSSTIW 180
QY	140	KIKSSKNSAANRELQKKLVKDKNKNRTKKKPKPPVVDVDEAGSLDNGDFKVTTPDST 199
Db	181	KIKSSKNSAANRELQKKLVKDKNKNRTKKKPKPPVVDVDEAGSLDNGDFKVTTPDST 240
QY	200	TQHNKSVTSPIITAKPINRPSLPNPSDTSKETSLSLVNKEITVETKETTNNKQSTDG 259
Db	241	TQHNKSVTSPIITAKPINRPSLPNPSDTSKETSLSLVNKEITVETKETTNNKQSTDG 300
QY	260	KEKTSIAKETQSIKTSKADLAPTSKVLAKPTKAEITTKGPALETTPEKPTPTPKPEPAS 319
Db	301	KEKTSIAKETQSIKTSKADLAPTSKVLAKPTKAEITTKGPALETTPEKPTPTPKPEPAS 360
QY	320	TTPEKPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 379
Db	361	TTPEKPTPTTKSAPTTPKPEAPTTTKSAPTTPKPEAPTTTKPEAPTTTKPEAPTTTKPE 420
QY	380	APTPTKSAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 439
Db	421	APTPTKSAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTPKPEAPTTPKPEAPTTPK 480
QY	440	EPAPTAKKAPTTPKPEAPTTPKPEAPTTTKKPSPTPKPEAPTTTKSAPTTTKPEAPT 499
Db	481	EPAPTAKKAPTTPKPEAPTTPKPEAPTTTKKPSPTPKPEAPTTTKSAPTTTKPEAPT 540
QY	500	TTKSAPTTPKPSPTTKPEAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTTKKPK 559
Db	541	TTKSAPTTPKPSPTTKPEAPTTPKPEAPTTPKKAPTTPKPEAPTTPKPEAPTTTKKPK 600
QY	560	APTAPKPEAPTTPKETAPTTPKLTTPPEKLAETTPPEKAPTTPEELAPTTPEEPTPTT 619
Db	601	APTAPKPEAPTTPKETAPTTPKLTTPPEKLAETTPPEKAPTTPEELAPTTPEEPTPTT 660
QY	620	PEEPAPTTPKAAADNTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTPKGTAPTTLKEP 679
Db	661	PEEPAPTTPKAAADNTPKPEAPTTPKPEAPTTPKETAPTTPKGTAPTTPKGTAPTTLKEP 720
QY	680	APTTPKKAPKELAPTTTKETSTSTSDKPAETTPKGTAPTTPKGPAPTTPKGPAPTTPKGP 739
Db	721	APTTPKKAPKELAPTTTKETSTSTSDKPAETTPKGTAPTTPKGPAPTTPKGPAPTTPKGP 780
QY	740	TAPTTLKBPAPTTPKKAPKELAPTTTKGPTSTTSDKPAETTPKGTAPTTPKGPAPTTPK 799
Db	781	TAPTTLKBPAPTTPKKAPKELAPTTTKGPTSTTSDKPAETTPKGTAPTTPKGPAPTTPK 840
QY	800	KPAPTTPETPTTSEVSTPTTTTKPTTIHKSPDSESTBELSAEPTPKALENSPKPEGVPT 859
Db	841	KPAPTTPETPTTSEVSTPTTTTKPTTIHKSPDSESTBELSAEPTPKALENSPKPEGVPT 900
QY	860	TKTPTAATKPEMTTAKDKTTERDLTPTTETTTAAKPKMTKETATTTKTTESKITATTTOV 919
Db	901	TKTPTAATKPEMTTAKDKTTERDLTPTTETTTAAKPKMTKETATTTKTTESKITATTTOV 960
QY	920	TSTTTQDTPPKITTLKITLAPKVTITTKITITTEIMNKPBEETAKPKDRATNSKATTPK 979
Db	961	TSTTTQDTPPKITTLKITLAPKVTITTKITITTEIMNKPBEETAKPKDRATNSKATTPK 1020
QY	980	POKPTKAPKKPTSTKPKTMPRVKPTTPTPRKMTSTMPELNPTSPRIAEAMLQTTTRPN 1039
Db	1021	POKPTKAPKKPTSTKPKTMPRVKPTTPTPRKMTSTMPELNPTSPRIAEAMLQTTTRPN 1080
QY	1040	QTPNSKLVEVNPKSEDAAGESTPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1099
Db	1081	QTPNSKLVEVNPKSEDAAGESTPHMLLRPHVFMPEVTPDMDYLPVNPQGIINPMLS 1140
QY	1100	DETNICNGKPVQGLTTLRNGTLVAFRGHYFWMLSFSPSPSPARRITEVWGIPSPIDTVFT 1159
Db	1141	DETNICNGKPVQGLTTLRNGTLVAFRGHYFWMLSFSPSPSPARRITEVWGIPSPIDTVFT 1200
QY	1160	RCNCEGKTFPFDKQSVWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY 1219
Db	1201	RCNCEGKTFPFDKQSVWRFTNDIKDAGYKPKPIFKGFGGLTGQIVAAALSTAKYKNWPESVY 1260
QY	1220	FFKRGGSIOQYIVKQBPVQKCPGRRPALNVPVVGEMTQVRRRRFERAIGPSQHTTIRIQY 1279
Db	1261	FFKRGGSIOQYIVKQBPVQKCPGRRPALNVPVVGEMTQVRRRRFERAIGPSQHTTIRIQY 1320
QY	1280	SPARLAYQDKGVLHNEVKVLSILWRGLPNVVTSAISLPNIRKPDGQDYDYAFSKQOYNNIDV 1339
Db	1321	SPARLAYQDKGVLHNEVKVLSILWRGLPNVVTSAISLPNIRKPDGQDYDYAFSKQOYNNIDV 1380
QY	1340	PSRTARAITRSGQTLSKVWYNCP 1363
Db	1381	PSRTARAITRSGQTLSKVWYNCP 1404
RESULT 4		
US-09-298-970A-1		
; Sequence 1, Application US/092989970A		
; Patent No. 6743774		
; GENERAL INFORMATION:		
; APPLICANT: Jay, Gregory D.		
; TITLE OF INVENTION: TRIBONECTINS		
; FILE REFERENCE: 21486-026		
; CURRENT APPLICATION NUMBER: US/09/298,970A		
; CURRENT FILING DATE: 2001-06-19		
; NUMBER OF SEQ ID NOS: 16		
; SOFTWARE: FastSeq for Windows Version 4.0		
; SEQ ID NO 1		
; LENGTH: 1404		
; TYPE: PRT		
; ORGANISM: Homo sapiens		
US-09-298-970A-1		
Query Match 99.8%; Score 7261.9; DB 4; Length 1404;		
Best Local Similarity 97.1%; Pred. No. 4.2e-196;		
Matches 1363; Conservative 0; Mismatches 0; Indels 41; Gaps 1;		
QY	1	MAWKTLPIYLILLLSVFVIQVSSQ-----25
Db	1	MAWKTLPIYLILLLSVFVIQVSSQDLSSCAGRCGEGYSRDATCNCYNCQHYMECCPDF 60
QY	26	-----ELSCCKGRCFESFERGREGDCDCAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 79



Db 61 KRVCTAELSCGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPSPSSKKAP 120  
Qy 80 PPSGASQTIKSTTKRSKPPNNKTKKVIIEEITEEHSVSENQESSSSSSSSSTIWI 139  
Db 121 PPSGASQTIKSTTKRSKPPNNKTKKVIIEEITEEHSVSENQESSSSSSSSSTIWI 180  
Qy 140 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDST 199  
Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDST 240  
Qy 200 TQHNKVSSTPKITTTAKINPRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDG 259  
Db 241 TQHNKVSSTPKITTTAKINPRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDG 300  
Qy 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 319  
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 360  
Qy 320 TTPKEPTTTPKEPASPTTKSAPTTTKSAPTTTPKEPATTTTPKEPATTTTPKEPAT 379  
Db 361 TTPKEPTTTPKEPASPTTKSAPTTTKSAPTTTPKEPATTTTPKEPATTTTPKEPAT 420  
Qy 380 APATTKSAPTTTPKEPATTTPKKAPTTTPKEPATTTTPKEPATTTTPKEPATTTPK 439  
Db 421 APATTKSAPTTTPKEPATTTPKKAPTTTPKEPATTTTPKEPATTTTPKEPATTTPK 480  
Qy 440 EPAPTAPKKDAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT 499  
Db 481 EPAPTAPKKDAPTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT 540  
Qy 500 TTKSAPTTTPKEPASPTTKKEPATTTTPKEPATTTTPKGPATTTTPKEPATTTTPK 559  
Db 541 TTKSAPTTTPKEPASPTTKKEPATTTTPKEPATTTTPKGPATTTTPKEPATTTTPK 600  
Qy 560 APATAPKEPATTTTPKEPATTTTPKGLTTPTPKGLAPTTTPKEPATTTTPBELAPTT 619  
Db 601 APATAPKEPATTTTPKEPATTTTPKGLTTPTPKGLAPTTTPKEPATTTTPBELAPTT 660  
Qy 620 PSEPATTTPKAAANPTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT 679  
Db 661 PSEPATTTPKAAANPTPKEPATTTTPKEPATTTTPKEPATTTTPKEPATTTTPKEPAT 720  
Qy 680 APATPKAPKELAPTTTKEPTSTTSKDPATTTTPKGTATTTTPKEPATTTTPKEPAT 739  
Db 721 APATPKAPKELAPTTTKEPTSTTSKDPATTTTPKGTATTTTPKEPATTTTPKEPAT 780  
Qy 740 TAPTTILKEPATTTPKKAPKELAPTTTKEPTSTTSKDPATTTTPKGTATTTTPKEPAT 799  
Db 781 TAPTTILKEPATTTPKKAPKELAPTTTKEPTSTTSKDPATTTTPKGTATTTTPKEPAT 840  
Qy 800 KPAPTTPPETPTTSEVSTPTTKEPTTTHKSPDESTPELSAEPPTPKALENSPKPGVPT 859  
Db 841 KPAPTTPPETPTTSEVSTPTTKEPTTTHKSPDESTPELSAEPPTPKALENSPKPGVPT 900  
Qy 860 TKTPAATKPEMTTAKOKTTERDLRTTPBTTTAAKPMKETAATTTTKEKTESKITATTTOV 919  
Db 901 TKTPAATKPEMTTAKOKTTERDLRTTPBTTTAAKPMKETAATTTTKEKTESKITATTTOV 960  
Qy 920 TSTTTQDTPPKITLTKTTILAPKVTTHKGTITTTTEIMNKPEETAKPKDRATNSKATTPK 979  
Db 961 TSTTTQDTPPKITLTKTTILAPKVTTHKGTITTTTEIMNKPEETAKPKDRATNSKATTPK 1020  
Qy 980 POKPTKAPKKPTSTKPKTMPVRPKPTTPTRKMTSTNPELNPTSRTAEMLOTTTRPN 1039  
Db 1021 POKPTKAPKKPTSTKPKTMPVRPKPTTPTRKMTSTNPELNPTSRTAEMLOTTTRPN 1080  
Qy 1040 QTPNSKLVENVNPKSEDAGAGETHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1099  
Db 1081 QTPNSKLVENVNPKSEDAGAGETHMLLRPHVFMPEVTPDMDYLPVFNQGIINPMLS 1140  
Qy 1100 DETNICNGKPVUDGLTTLRNGTLVAFRGHYFWMLSPPSPARRITEVWGIPSPIDTFT 1159

Db 1141 DETNICNGKPVUDGLTTLRNGTLVAFRGHYFWMLSPPSPARRITEVWGIPSPIDTFT 1200  
Qy 1160 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKDFKGFGLTGQIYAALSTAKYKNWPESVY 1219  
Db 1201 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPKDFKGFGLTGQIYAALSTAKYKNWPESVY 1260  
Qy 1220 PFKEGGSIQYIYKQEPVQKCPGRRPALNTPVYGMTQVRRRRFERAIGPSQTHIRIQY 1279  
Db 1261 PFKEGGSIQYIYKQEPVQKCPGRRPALNTPVYGMTQVRRRRFERAIGPSQTHIRIQY 1320  
Qy 1280 SPARLAYQDKGVLHNEVKVSLMRGLNVTSAISLNPBKPDGYDYAASKDQYYNIDV 1339  
Db 1321 SPARLAYQDKGVLHNEVKVSLMRGLNVTSAISLNPBKPDGYDYAASKDQYYNIDV 1380  
Qy 1340 PSRTARAITTRSGQTLISKVWYNCP 1363  
Db 1381 PSRTARAITTRSGQTLISKVWYNCP 1404

RESULT 5  
US-10-164-595-78  
; Sequence 78, Application US/10164595  
; Patent No. 6657054  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies, Inc  
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
; FILE REFERENCE: IU 103 R1  
; CURRENT APPLICATION NUMBER: US/10/164,595  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 78  
; LENGTH: 1404  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-595-78

Query Match 99.4%; Score 7232.9; DB 4; Length 1404;  
Best Local Similarity 96.8%; Pred. No. 2.7e-195;  
Matches 1359; Conservative 0; Mismatches 4; Indels 41; Gaps 1;

Qy 1 MAWKTLPIYLLLLSVFVIQQVSSQ----- 25  
Db 1 MAWKTLPIYLLLLSVFVIQQVSSQDLSSCAGRCGEGYSRDATCNDCYNCOHYMECCPDF 60  
Qy 26 -----ELSCGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPSPSSKKAP 79  
Db 61 KRVCTAELSCGRCFESFERGECDDAQQCKYDKCCPDYESFCAEVHNPSPSSKKAP 120  
Qy 80 PPSGASQTIKSTTKRSKPPNNKTKKVIIEEITEEHSVSENQESSSSSSSSSTIWI 139  
Db 121 PPSGASQTIKSTTKRSKPPNNKTKKVIIEEITEEHSVSENQESSSSSSSSSTIWI 180  
Qy 140 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDST 199  
Db 181 KIKSSKNSAANRELQKKLVKDNKNRTKKKPTPKPPVVDAGSLDNGDFKVTTPDST 240  
Qy 200 TQHNKVSSTPKITTTAKINPRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDG 259  
Db 241 TQHNKVSSTPKITTTAKINPRPSLPNSDTSKETSITVKNKETTVEKTTTNNKQTSIDG 300  
Qy 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 319  
Db 301 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTTPKEPTTTPKEPAS 360  
Qy 320 TTPKEPTTTPKEPASPTTKSAPTTTKSAPTTTPKEPATTTTPKEPATTTTPKEPAT 379  
Db 361 TTPKEPTTTPKEPASPTTKSAPTTTKSAPTTTPKEPATTTTPKEPATTTTPKEPAT 420  
Qy 380 APATTKSAPTTTPKEPATTTPKKAPTTTPKEPATTTTPKEPATTTTPKEPATTTPK 439  
Db 421 APATTKSAPTTTPKEPATTTPKKAPTTTPKEPATTTTPKEPATTTTPKEPATTTPK 480



```
QY 361 KEPAATTTKEPAATTTKEPAATTTKSAATTTKEPAATTTKEPAATTTKEPAATTTKEPTP 420
Db 318 KEPAATTTKEPAATTTKEPAATTTKSAATTTKEPAATTTKEPAATTTKEPAATTTKEPTP 377
QY 421 TTKEPAATTTKEPAATTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTTKEPSTTKE 480
Db 378 TTKEPAATTTKEPAATTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTTKEPSTTKE 437
QY 481 PAATTTKSAATTTKEPAATTTKSAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTT 540
Db 438 PAATTTKSAATTTKEPAATTTKSAATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTT 497
QY 541 KEPAATTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTT 600
Db 498 KEPAATTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTTKEPAATTTKEPAATTT 557
QY 601 PTPEELAPTTPEEPPTTTPEEPATTTKAAAPNTKPEPAATTTKEPAATTTKEPAATTT 660
Db 558 PTPEELAPTTPEEPPTTTPEEPATTTKAAAPNTKPEPAATTTKEPAATTTKEPAATTT 617
QY 661 KETAPTTKGTAPTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTTKEPAATTT 720
Db 618 KETAPTTKGTAPTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTTKEPAATTT 677
QY 721 KEPAATTTKEPAATTTKGTAPTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTT 780
Db 678 KEPAATTTKEPAATTTKGTAPTTKEPAATTTKAPKAPATTTKEPAATTTKEPAATTT 737
QY 781 TPKETAPTTKEPAATTTKEPAATTTPEEPPTTSEVSTPTTKEPTTIHKSPDESTPELS 840
Db 738 TPKETAPTTKEPAATTTKEPAATTTPEEPPTTSEVSTPTTKEPTTIHKSPDESTPELS 797
QY 841 AEPTPKALENSKEPGVPTTKTTPAATKPEMTTAKDKTTERDLRTTPELTAAAPKMTKET 900
Db 798 AEPTPKALENSKEPGVPTTKTTPAATKPEMTTAKDKTTERDLRTTPELTAAAPKMTKET 857
QY 901 ATTETKTESKITATTQVTSSTTQDTTPPKITLTKITTLAPKVTITTKITITTEIMNKP 960
Db 858 ATTETKTESKITATTQVTSSTTQDTTPPKITLTKITTLAPKVTITTKITITTEIMNKP 917
QY 961 EETAKPKDRATNSKATTPKPKETKAPKPTSTKPKTMRVPRKPTTTPRKMSTMP 1020
Db 918 EETAKPKDRATNSKATTPKPKETKAPKPTSTKPKTMRVPRKPTTTPRKMSTMP 977
QY 1021 LNPTSRIAEAMLQTTTRPQNTNSKLVNPNKSEDAAGAEGETPHMLLRPHVPMPEVTPD 1080
Db 978 LNPTSRIAEAMLQTTTRPQNTNSKLVNPNKSEDAAGAEGETPHMLLRPHVPMPEVTPD 1037
QY 1081 MDYLRVPVNOGIIINPMLSDETNI CNKPKVDGLTTLRNGTLVAFRGHYFWMLSPFSPSP 1140
Db 1038 MDYLRVPVNOGIIINPMLSDETNI CNKPKVDGLTTLRNGTLVAFRGHYFWMLSPFSPSP 1097
QY 1141 ARRIITEVWGIPIPSIDTFTRCNCEGKTFEFKDSQYWRFTNDIKDAGYKPIPKFGGGLTG 1200
Db 1098 ARRIITEVWGIPIPSIDTFTRCNCEGKTFEFKDSQYWRFTNDIKDAGYKPIPKFGGGLTG 1157
QY 1201 QIVAAALSTAKYKNWPESVYFFKRGGSIQYIYKQEPVQKCGRRPALNVPVGMTOVER 1260
Db 1158 QIVAAALSTAKYKNWPESVYFFKRGGSIQYIYKQEPVQKCGRRPALNVPVGMTOVER 1217
QY 1261 RFPERAIGPSQTHTRI QIVSPARLAYQDKGVHLNEVKVSI LWRGLFNVVTSAISLPNIRK 1320
Db 1218 RFPERAIGPSQTHTRI QIVSPARLAYQDKGVHLNEVKVSI LWRGLFNVVTSAISLPNIRK 1277
QY 1321 PDGYDYAFSKDQYNNIDVPSTABAITTRSQTLISKVWYNCP 1363
Db 1278 PDGYDYAFSKDQYNNIDVPSTABAITTRSQTLISKVWYNCP 1320
```

RESULT 7

US-07-757-022B-60

; Sequence 60, Application US/07757022B

```
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Gesner, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1320 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-60
```

Query Match 97.0%; Score 7058.7; DB 4; Length 1320;  
Best Local Similarity 96.8%; Pred. No. 1.9e-190;  
Matches 1320; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

```
QY 1 MAWKLTPIYLLILLLVFVIQQVSSOELSCKGRCPESFERGRCDCDAQCKKYDKCCPDYE 60
Db 1 MAWKLTPIYLLILLLVFVIQQVSSOELSCKGRCPESFERGRCDCDAQCKKYDKCCPDYE 60
QY 61 SFCAEVHNPTSPSSKKAAPPSPGASQTIKSTTKSPKPNKKTKKVIKESBEITEHSVS 120
Db 61 SFCAEVHNPTSPSSKKAAPPSPGASQTIKSTTKSPKPNKKTKKVIKESBEITE 115
QY 121 ENQESSSSSSSSSSSTIWKIKSSKNAAARELQKLVKDKNKKRTKKKTKPKPPVYDE 180
Db 116 -----VKDNKKRTKKKTKPKPPVYDE 137
QY 181 AGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRLPNSDTSKETSITVNKE 240
Db 138 AGSGLDNGDFKVTTPDTSTTQHNKVSTSPKITTAKPINPRLPNSDTSKETSITVNKE 197
QY 241 TTVTETKTTTNTKQSTIDGKEKTTSAKETQSIKTSKADLAPTSKVLAKPETTKG 300
```

Db	198	TTVETKETTNNKQTSDCKEKTTGAKETQSIKTSADLAPTQVLAKPTPKASTTNG	257
Qy	301	PALTTTPKEPTPTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP	360
Db	258	PALTTTPKEPTPTTPKEPASTTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP	317
Qy	361	KEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPTTP	420
Db	318	KEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKEPTTP	377
Qy	421	TTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE	480
Db	378	TTTPKEPAPTTPKEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKE	437
Qy	481	PAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP	540
Db	438	PAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTP	497
Qy	541	KEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP	600
Db	498	KEPAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP	557
Qy	601	PTTPELAPTTPKEPTPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP	660
Db	558	PTTPELAPTTPKEPTPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP	617
Qy	661	KETAPTTPKGTAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKGTAPT	720
Db	618	KETAPTTPKGTAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKEPAPTTPKGTAPT	677
Qy	721	PKSPAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKGTAPT	780
Db	678	PKSPAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKGTAPT	737
Qy	781	TPKETAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKGTAPT	840
Db	738	TPKETAPTTPKEPAPTTPKGTAPTTPKEPAPTTPKSAPTTPKEPAPTTPKSAPTTPKGTAPT	797
Qy	841	AEPTPKALENSPKPGVPTTKPAATKPEMTTAKDKTTERDLRTPPETTTAAPKMTKET	900
Db	798	AEPTPKALENSPKPGVPTTKPAATKPEMTTAKDKTTERDLRTPPETTTAAPKMTKET	857
Qy	901	ATTTEKTESKITATTTQVSTTTQDTPPKITLTKTTLAPKVTTKKTLITTEIMNKP	960
Db	858	ATTTEKTESKITATTTQVSTTTQDTPPKITLTKTTLAPKVTTKKTLITTEIMNKP	917
Qy	961	BETAKPKDRATNSKATTPKOKPTKAPKPTSTKPKTTPMPVRKPTTPPKMTSTWPE	1020
Db	918	BETAKPKDRATNSKATTPKOKPTKAPKPTSTKPKTTPMPVRKPTTPPKMTSTWPE	977
Qy	1021	LNPTSRIAEAMLOTTTRNCPQNSKLEVNPKSBDAGAGETPHMLLRPHVFMPEVTPD	1080
Db	978	LNPTSRIAEAMLOTTTRNCPQNSKLEVNPKSBDAGAGETPHMLLRPHVFMPEVTPD	1037
Qy	1081	MDYLPRVNPQIILNPMLSDETNCNGKPVGLTTLRNGTLVAFRGHVFWMLSFSPSPSP	1140
Db	1038	MDYLPRVNPQIILNPMLSDETNCNGKPVGLTTLRNGTLVAFRGHVFWMLSFSPSPSP	1097
Qy	1141	ARRITEVWGIPIPDITVTRCNCEGKTFEFDKSDQYWRFTNDIKDAGYKPKPIFKGFGGLTG	1200
Db	1098	ARRITEVWGIPIPDITVTRCNCEGKTFEFDKSDQYWRFTNDIKDAGYKPKPIFKGFGGLTG	1157
Qy	1201	QIVAAALSTAKYKNWPESVYFFKRGSGIOQYIYKQEPVOKCPGRRPALNYPVYGMTQVRR	1260
Db	1158	QIVAAALSTAKYKNWPESVYFFKRGSGIOQYIYKQEPVOKCPGRRPALNYPVYGMTQVRR	1217
Qy	1261	RRFERAIGESQHTTIRIQYSPARLAYQDKGVHLNPKVKSILLWRGLPNVVTSAISLPNIRK	1320
Db	1218	RRFERAIGESQHTTIRIQYSPARLAYQDKGVHLNPKVKSILLWRGLPNVVTSAISLPNIRK	1277
Qy	1321	PDGYDYAFSKQDYNNIDVPSRTARAITTRSGQTLTKSVWYNCP	1363
Db	1278	PDGYDYAFSKQDYNNIDVPSRTARAITTRSGQTLTKSVWYNCP	1320

RESULT 8  
US-07-757-022B-40  
; Sequence 40, Application US/07757022B  
; Patent No. 6433142  
; GENERAL INFORMATION:  
; APPLICANT: Geener, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseerr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1361 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-757-022B-40

Query Match 96.8%; Score 7044.6; DB 4; Length 1361;  
Best Local Similarity 94.0%; Pred. No. 5.1e-190;  
Matches 1320; Conservative 0; Mismatches 0; Indels 84; Gaps 2;

Qy	1	MAWKTLPIYLLLLSVFVIQVSSQ-----	25
Db	1	MAWKTLPIYLLLLSVFVIQVSSQ-----	60
Qy	26	-----ELSCGRCFESPERGECDCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP	79
Db	61	KVCTAELSCGRCFESPERGECDCDAQCKKYDKCCPDYESFCAEVHNPTSPSSKKAP	120
Qy	80	PPSGASQTIKSTTKRSPKPPNKKTKVIESEETEEHSVSENQSSSSSSSSSSSTIW	139
Db	121	PPSGASQTIKSTTKRSPKPPNKKTKVIESEETEEHSVSENQSSSSSSSSSSSTIW	156
Qy	140	KIKSSKNSAANRELOKKLKVDKNKNTKKKPTPKPPVVDEAGSLDNGDFKVTTPDTST	199

Db 157 -----VKDNKKNTKKKPTPKPPVDEAGSLONGDFKVTTPDST 197  
QY 200 TQHNKYSTSKITAKPINSRPSLPNNSDTSKESLTVNKETTVEKTTTNNKQSTDG 259  
Db 198 TQHNKYSTSKITAKPINSRPSLPNNSDTSKESLTVNKETTVEKTTTNNKQSTDG 257  
QY 260 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 319  
Db 258 KEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKGPALTTKPEPTTTPKEPAS 317  
QY 320 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKGPAPTTTPKEPAPTTTKEP 379  
Db 318 TTPKEPTPTTIKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKGPAPTTTPKEPAPTTTKEP 377  
QY 380 APPTTKSAPTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 439  
Db 378 APPTTKSAPTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 437  
QY 440 EPAPTAPKAPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 499  
Db 438 EPAPTAPKAPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 497  
QY 500 TTKSAPTTPKEPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 559  
Db 498 TTKSAPTTPKEPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 557  
QY 560 APAPTAPKEPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 619  
Db 558 APAPTAPKEPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 617  
QY 620 PEEPAPTTPKAAAPNTPKPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTT 679  
Db 618 PEEPAPTTPKAAAPNTPKPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTT 677  
QY 680 APPTTKKAPKELAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 739  
Db 678 APPTTKKAPKELAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 737  
QY 740 TAPPTLKAPAPTTTPKAPKELAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTT 799  
Db 738 TAPPTLKAPAPTTTPKAPKELAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTT 797  
QY 800 KPAPTTPPTPTTSEVSTPTTKBPTTIHKSPDESTPELSAETPKALENSPKBGPVT 859  
Db 798 KPAPTTPPTPTTSEVSTPTTKBPTTIHKSPDESTPELSAETPKALENSPKBGPVT 857  
QY 860 TKTPAATKPEMTTAKDKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTITQV 919  
Db 858 TKTPAATKPEMTTAKDKTTERDLRTTPTTTTAAAPKMTKETATTTTEKTTESKITATTITQV 917  
QY 920 TSTTTTQDTTPPKITTLTKTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK 979  
Db 918 TSTTTTQDTTPPKITTLTKTTLAPKVTITTKTITTTIMNKPEETAKPKDRATNSKATTPK 977  
QY 980 POKPTKAPKKPTSTKKPKTTPMVRPKTTPTPRKMSTMPBLNPTSRIAEAMLOTTTTPN 1039  
Db 978 POKPTKAPKKPTSTKKPKTTPMVRPKTTPTPRKMSTMPBLNPTSRIAEAMLOTTTTPN 1037  
QY 1040 QTPNSKLVEVNPKSDAGAGETHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1099  
Db 1038 QTPNSKLVEVNPKSDAGAGETHMLLRPHVFMPEVTPDMDYLPRVFNQGIINPMLS 1097  
QY 1100 DETNI CNKGPVDGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRITEVWGIPSPDITVFT 1159  
Db 1098 DETNI CNKGPVDGLTTLRNGTLVAPRGHYFWMLSPPSPSPARRITEVWGIPSPDITVFT 1157  
QY 1160 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIIVAALSTAKYKNWPESVY 1219  
Db 1158 RCNCEGKTFPFKDSQYWRFTNDIKDAGYKPIFKGFGGLTGOIIVAALSTAKYKNWPESVY 1217  
QY 1220 FFKRGGSIQOYIYKQEPVQKCGRRPALNYPVYGMTQVRRRRRFPRAIGPSQTHIRIQY 1279

Db 1218 FFKRGGSIQOYIYKQEPVQKCGRRPALNYPVYGMTQVRRRRRFPRAIGPSQTHIRIQY 1277  
QY 1280 SPARLAYQDKGVLHNEVKVSIILMRGLPNVVTSAISLFINIRKPDGYDYIYAFSKOYYNIDV 1339  
Db 1278 SPARLAYQDKGVLHNEVKVSIILMRGLPNVVTSAISLFINIRKPDGYDYIYAFSKOYYNIDV 1337  
QY 1340 PSRTARAITTSGOTLSKWNYNCP 1363  
Db 1338 PSRTARAITTSGOTLSKWNYNCP 1361  
RESULT 9  
US-10-164-595-58  
; Sequence 58, Application US/10164595  
; Patent No. 6657054  
; GENERAL INFORMATION:  
; APPLICANT: Origene Technologies, Inc  
; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides  
; FILE REFERENCE: IU 103 P1  
; CURRENT APPLICATION NUMBER: US/10/164,595  
; CURRENT FILING DATE: 2002-06-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 58  
; LENGTH: 1320  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-164-595-58  
Query Match 96.8%; Score 7043.7; DB 4; Length 1320;  
Best Local Similarity 96.6%; Pred. No. 5.1e-190;  
Matches 1317; Conservative 0; Mismatches 3; Indels 43; Gaps 1;  
QY 1 MAWKTLPIYLLLLSVFVIQQVSSOBLSCGKCFESFERGREGCDCAQCKKYDKKCCPDYE 60  
Db 1 MAWKTLPIYLLLLSVFVIQQVSSOBLSCGKCFESFERGREGCDCAQCKKYDKKCCPDYE 60  
QY 61 SFCAEVHNPSTSPSSKAPPPSGASQTITKSTTRGPKPPNKKTKKVIIESEITEHSVS 120  
Db 61 SFCAEVHNPSTSPSSKAPPPSGASQTITKSTTRGPKPPNKKTKKVIIESEITE 115  
QY 121 ENQESSSSSSSSSTIWKIKSSNNAANRELQKKLVKDNKKARTKKKPTPKPPVDE 180  
Db 116 -----VKDNKKARTKKKPTPKPPVDE 137  
QY 181 AGSGLDNGDFKVTTPDSTTQHNKYSTSKITAKPINSRPSLPNNSDTSKESLTVNKE 240  
Db 138 AGSGLDNGDFKVTTPDSTTQHNKYSTSKITAKPINSRPSLPNNSDTSKESLTVNKE 197  
QY 241 TTVETKETTTNNKQSTDGKEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKG 300  
Db 198 TTVETKETTTNNKQSTDGKEKTTSAKETQSIEKTSKADLAPTSKVLAKPTPKAETTTKG 257  
QY 301 PALTTTPKEPTPTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 360  
Db 258 PALTTTPKEPTPTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTT 317  
QY 361 KEPAFTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTKGPAPTT 420  
Db 318 KEPAFTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTKGPAPTT 377  
QY 421 TTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 480  
Db 378 TTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTPKEPAPTTTKGPAPTTTKGPAPTTTKGPAPTT 437  
QY 481 PAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTT 540  
Db 438 PAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTTTKSAPTTPKEPAPTT 497  
QY 541 KEPAFTTPKEPAPTTTKKAPAPTTAPKAPAPTTAPKAPAPTTAPKAPAPTTAPKAPAPTTAPKAPAPTT 600  
Db 498 KEPAFTTPKEPAPTTTKKAPAPTTAPKAPAPTTAPKAPAPTTAPKAPAPTTAPKAPAPTTAPKAPAPTT 557

```
QY 601 PTTPELAPTTPEPTTTPPEPAPTTPKAAAPNTKEPAPTTKEPAPTTKEPAPTT 660
Db 558 PTTPELAPTTPEPTTTPPEPAPTTPKAAAPNTKEPAPTTKEPAPTTKEPAPTT 617
QY 661 KETAPTTKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTISKAPATTPKGTAPTT 720
Db 618 KETAPTTKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTTCDKAPATTPKGTAPTT 677
QY 721 PKEPAPTTKEPAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTISKAPATT 780
Db 678 PKEPAPTTKEPAPTTPKGTAPTTLKEPAPTTPKKAPKELAPTTTKEPTSTISKAPATT 737
QY 781 TPKEPAPTTKEPAPTTPKKAPKELAPTTTKEPTSTISKAPATTTKEPTTTHKSDESTEP 840
Db 738 TPKEPAPTTKEPAPTTPKKAPKELAPTTTKEPTSTISKAPATTTKEPTTTHKSDESTEP 797
QY 841 ABPTPKALNSKPEKGPVPTTKPAATKPEMTTITAKDKTTTERDLRTTETTTAAPKMTKET 900
Db 798 ABPTPKALNSKPEKGPVPTTKPAATKPEMTTITAKDKTTTERDLRTTETTTAAPKMTKET 857
QY 901 ATTTEKTSKGTATTTQVTSITTTQDTPPKITTTTLKTTTLAPKVTTKKTTTITTEIMNKP 960
Db 858 ATTTEKTSKGTATTTQVTSITTTQDTPPKITTTTLKTTTLAPKVTTKKTTTITTEIMNKP 917
QY 961 BETAKPKDRATNSKATTPKPKQKTPAKPKPTSTKPKKTWPRVKPKTTPTRKMTSTMPD 1020
Db 918 BETAKPKDRATNSKATTPKPKQKTPAKPKPTSTKPKKTWPRVKPKTTPTRKMTSTMPD 977
QY 1021 LNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1080
Db 978 LNPTSRIAEAMLTQTTTRPNQTPNSKLVEVNPKSEDAGGAGETPHMLLRPHVFMPEVTPD 1037
QY 1081 MDYLPVPNGQIILNPMLSDETINICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSP 1140
Db 1038 MDYLPVPNGQIILNPMLSDETINICNGKPDVGLTTLRNGTLVAFRGHYFWMLSPPFPSP 1097
QY 1141 ARSITEVWGPSPIDVTFRCNCEGKTFEPKDSQYWRFTNDIKDAGVPKPIFKGFGGLTG 1200
Db 1098 ARSITEVWGPSPIDVTFRCNCEGKTFEPKDSQYWRFTNDIKDAGVPKPIFKGFGGLTG 1157
QY 1201 QIVAAALSTAKYKNWPESVVFYFKRGSGSIQQYIYKQEPVQKCPGRRPALNYPVYGMTQVRR 1260
Db 1158 QIVAAALSTAKYKNWPESVVFYFKRGSGSIQQYIYKQEPVQKCPGRRPALNYPVYGMTQVRR 1217
QY 1261 RRFERAIGPSQTHIRIQYSPARLAYQDKGVILHNEVKVSIILWGLPNVVTSAISLPNIRK 1320
Db 1218 RRFERAIGPSQTHIRIQYSPARLAYQDKGVILHNEVKVSIILWGLPNVVTSAISLPNIRK 1277
QY 1321 PDGVYVAFSKDQYVNIIDVPSRTARAITTRSGQTLKSVWYNCP 1363
Db 1278 PDGVYVAFSKDQYVNIIDVPSRTARAITTRSGQTLKSVWYNCP 1320
```

```
RESULT 10
US-07-757-022B-142
; Sequence 142, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Bridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 142:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1313 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-142
```

```
Query Match 96.2%; Score 7002; DB 4; Length 1313;
Best Local Similarity 96.3%; Pred. No. 7.5e-189;
Matches 1313; Conservative 0; Mismatches 0; Indels 50; Gaps 1;
```

```
QY 1 MAWKTPYVLLLSLVFVIQQVSSQELSKGRCFESFERGRCDAQCKKYDKCCPDYE 60
Db 1 MAWKTPYVLLLSLVFVIQQVSSQELSKGRCFESFERGRCDAQCKKYDKCCPDYE 60
QY 61 SFCAEVHNETSPSSKKAPPPSGASQTIKSTTKRSPKPNKTKTKVIESEETEHSHVS 120
Db 61 SFCA-----EESHVS 70
QY 121 ENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKNRTKKKTPKPPVVDE 180
Db 71 ENQESSSSSSSSSSSTIWKIKSSKNSAANRELQKLVKDKNKNRTKKKTPKPPVVDE 130
QY 181 AGSGLDNGDPKVTTPDTSTTQHNKVSTSPKITTAKPINRPPSLPPNSDTSKETSILTWNKE 240
Db 131 AGSGLDNGDPKVTTPDTSTTQHNKVSTSPKITTAKPINRPPSLPPNSDTSKETSILTWNKE 190
QY 241 TTVETKETTTNNKQSTDGKEKTTSAKETQSIEKTSADLAPTSKVLAKTPKAEETTKG 300
Db 191 TTVETKETTTNNKQSTDGKEKTTSAKETQSIEKTSADLAPTSKVLAKTPKAEETTKG 250
QY 301 PALTTPEPTTTPKEPASTTKEPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTT 360
Db 251 PALTTPEPTTTPKEPASTTKEPTTTIKSAPTTTKEPAPTTTKSAPTTTKEPAPTTT 310
QY 361 KEPAPTTPEPAPTTTKEPAPTTTKSAPTTTKEPAPTTTKKAPATTTPKEPAPTTTKEPTP 420
Db 311 KEPAPTTPEPAPTTTKEPAPTTTKSAPTTTKEPAPTTTKKAPATTTPKEPAPTTTKEPTP 370
QY 421 TTPKEPAPTTTKEPAPTTTKEPAPTTAPKPKAPATTTPKEPAPTTTKEPAPTTTKEPSPTTKE 480
Db 371 TTPKEPAPTTTKEPAPTTTKEPAPTTAPKPKAPATTTPKEPAPTTTKEPAPTTTKEPSPTTKE 430
QY 481 PAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 540
Db 431 PAPTTTTSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 490
```

```
QY 541 KEPAATTPKBPATTTTKKPAATAPKEPATTPKETAATTPKKLPTTPEKLAPTPEKPA 600
Db 491 KEPAATTPKBPATTTTKKPAATAPKEPATTPKETAATTPKKLPTTPEKLAPTPEKPA 550
QY 601 PTTPEELAPTTPPEPTTTPPEPATTPKAAAPNTKPEPATTPKBPATTPKBPATTP 660
Db 551 PTTPEELAPTTPPEPTTTPPEPATTPKAAAPNTKPEPATTPKBPATTPKBPATTP 610
QY 661 KETAPTTPKGPATTLKEPATTPKPAKELATTTKEPTSTTSKPAATTPKGTAPT 720
Db 611 KETAPTTPKGPATTLKEPATTPKPAKELATTTKEPTSTTSKPAATTPKGTAPT 670
QY 721 PKEPATTPKBPATTPKGTAPTTLKEPATTPKPAKELATTTKGTSTTSKPAAT 780
Db 671 PKEPATTPKBPATTPKGTAPTTLKEPATTPKPAKELATTTKGTSTTSKPAAT 730
QY 781 TPKEATTPKBPATTPKGPATTPETPTPTTSEVSTPTTKETPTTIHKSPDESTPELS 840
Db 731 TPKEATTPKBPATTPKGPATTPETPTPTTSEVSTPTTKETPTTIHKSPDESTPELS 790
QY 841 AEPTPKALENSPKBGPVPTTKPAATKPEMTTAKDKTTERDLRTTETTTAAPKMTKET 900
Db 791 AEPTPKALENSPKBGPVPTTKPAATKPEMTTAKDKTTERDLRTTETTTAAPKMTKET 850
QY 901 ATTTEKTTESKITATTTQVSTTTQDTPPKKITTLKTTILAPKVTTTKITTTIEMNKP 960
Db 851 ATTTEKTTESKITATTTQVSTTTQDTPPKKITTLKTTILAPKVTTTKITTTIEMNKP 910
QY 961 EETAKPKDRATNSKATTPKQKPTKAPKPTSTTKKPTMPRVRKPTTTPRKMSTMPPE 1020
Db 911 EETAKPKDRATNSKATTPKQKPTKAPKPTSTTKKPTMPRVRKPTTTPRKMSTMPPE 970
QY 1021 LNPTSLAEAMLQTTTRENQTPNSKLVNPKSEDAGAGETPHMLLRHHVFMPEVTPD 1080
Db 971 LNPTSLAEAMLQTTTRENQTPNSKLVNPKSEDAGAGETPHMLLRHHVFMPEVTPD 1030
QY 1081 MDYLPVPVNOGIIINPMLSDETNICNGKPDVDTLRLNGTLVAPRGHYFMWLSPEPSP 1140
Db 1031 MDYLPVPVNOGIIINPMLSDETNICNGKPDVDTLRLNGTLVAPRGHYFMWLSPEPSP 1090
QY 1141 ARRITEWGPISPIDVTRCNCCEGKTFPPKDSQYWRFTNDIKDAGYKPIFKGFGGLTG 1200
Db 1091 ARRITEWGPISPIDVTRCNCCEGKTFPPKDSQYWRFTNDIKDAGYKPIFKGFGGLTG 1150
QY 1201 QIVAAALSTAKYKNWPESVVFYFKGGGSIQYIYKQEPVQKCPGRPALNPVYVGMTOVRR 1260
Db 1151 QIVAAALSTAKYKNWPESVVFYFKGGGSIQYIYKQEPVQKCPGRPALNPVYVGMTOVRR 1210
QY 1261 RRFERAIGPSQTHIRIQYSPARLAYQDKGVHLHNEVKVSIILWGLPNVVTSAISLBNIRK 1320
Db 1211 RRFERAIGPSQTHIRIQYSPARLAYQDKGVHLHNEVKVSIILWGLPNVVTSAISLBNIRK 1270
QY 1321 PDGYDYAFSKQYNYNDVPSRTARAITRRSGOTLSKWYNCP 1363
Db 1271 PDGYDYAFSKQYNYNDVPSRTARAITRRSGOTLSKWYNCP 1313
```

RESULT 11

```
US-07-757-022B-48
; Sequence 48, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
```

```
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1354 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-48
```

```
Query Match 96.0%; Score 6987.9; DB 4; Length 1354;
Best Local Similarity 93.5%; Pred. No. 2e-188; 0; Mismatches 0; Indels 91; Gaps 2;
Matches 1313; Conservative 0;

QY 1 MAWKTLPIVLLLLLVFVIOQVSSQ----- 25
Db 1 MAWKTLPIVLLLLLVFVIOQVSSQDLSSCAGRCGRGYSRDATCNDYNCQHYMECCPDF 60
QY 26 -----ELSCGRCFESFERGECDCDAOCKKYDKCCPDYESFCAEVHNPTSPSSKKAP 79
Db 61 KRVCTRAELSCGRCFESFERGECDCDAOCKKYDKCCPDYESFCA----- 105
QY 80 PPSGASQTIKSTTKRSPKPPNKKTKKVIETEEHSVSENQBSSESSSSSSSSSSSTI 139
Db 106 -----EEHSVSENQBSSESSSSSSSSSSSTI 130
QY 140 KIKSKNSAANRELOKLVKONKNRTKKKTPPPVVDVDEAGSLDNGDFKVTTPDTST 199
Db 131 KIKSKNSAANRELOKLVKONKNRTKKKTPPPVVDVDEAGSLDNGDFKVTTPDTST 190
QY 200 TOHNKVSTSPKLTITAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTTNKQSTDG 259
Db 191 TOHNKVSTSPKLTITAKPINRPSLPNSDTSKETSITVKNKETTVEKTTTTNKQSTDG 250
QY 260 KEKTTSAKETOSIEKTSARDLAPTSKVLAKPTPKAETTTKGPALTTPKETPTTTPKEPAS 319
Db 251 KEKTTSAKETOSIEKTSARDLAPTSKVLAKPTPKAETTTKGPALTTPKETPTTTPKEPAS 310
QY 320 TTPKEPTPTTTKSAPTTPKBPATTTKSAPTTPKBPATTTKBPATTTKBPATTTKBP 379
Db 311 TTPKEPTPTTTKSAPTTPKBPATTTKSAPTTPKBPATTTKBPATTTKBPATTTKBP 370
QY 380 APTTTKSAPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP 439
Db 380 APTTTKSAPTTPKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBPATTTKBP 439
```







QY 241 TTVETKEITTTNKQSTGDKKETTSAKETOSIEKTSKDLAPTSKVLAKPTPKAETTTKG 300  
Db 148 TTVETKEITTTNKQSTGDKKETTSAKETOSIEKTSKDLAPTSKVLAKPTPKAETTTKG 207  
QY 301 PALTTPKBETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 360  
Db 208 PALTTPKBETPTTPKEPASTTPKEPTPTTIKSAPTTPKEPAPTTPKSAPTTPKEPAPTTP 267  
QY 361 KBPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 420  
Db 268 KBPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 327  
QY 421 TTPKEPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 480  
Db 328 TTPKEPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBP 387  
QY 481 PAPTTTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 540  
Db 388 PAPTTTKSAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 447  
QY 541 KBPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 600  
Db 448 KBPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 507  
QY 601 PTTPEELAPTTPPEPTTPPEEAPTTPKAAANTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 660  
Db 508 PTTPEELAPTTPPEPTTPPEEAPTTPKAAANTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 567  
QY 661 KETAPTTPKGTAPTTLKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 720  
Db 568 KETAPTTPKGTAPTTLKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 627  
QY 721 PKBPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 780  
Db 628 PKBPAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 687  
QY 781 TPKETAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 840  
Db 688 TPKETAPTTPKEPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTPKBPAPTTP 747  
QY 841 AEPTPKALENSKPEPGVPTTKPAATKPEMTTTAKDKITBERDLRTTTPETTTAAPKMTKET 900  
Db 748 AEPTPKALENSKPEPGVPTTKPAATKPEMTTTAKDKITBERDLRTTTPETTTAAPKMTKET 807  
QY 901 ATTEKTESKITATTTQVSTTTQDTPPKIITLTKTLAPKVTITTKIITITTEIMNKP 960  
Db 808 ATTEKTESKITATTTQVSTTTQDTPPKIITLTKTLAPKVTITTKIITITTEIMNKP 867  
QY 961 EETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPT 1020  
Db 868 EETAKPKDRATNSKATTPKPKPTKAPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPTSTKPKPT 927  
QY 1021 LNPTSRIAEAMLTQTRPNQTPNSKLVENPKSBDAGAGETPHMLLRPHVFMVEVTPD 1080  
Db 928 LNPTSRIAEAMLTQTRPNQTPNSKLVENPKSBDAGAGETPHMLLRPHVFMVEVTPD 987  
QY 1081 MDYLPVFNQIIINPMLSDETNINCGRPVDGLTTLNGLTLAVARGHYFWMFLSPSPSPSP 1140  
Db 988 MDYLPVFNQIIINPMLSDETNINCGRPVDGLTTLNGLTLAVARGHYFWMFLSPSPSPSP 1047  
QY 1141 ARRITEVWGIPSPIDTVPTRCNCEGKTPFFKDSQVWRFTNDIKDAGYKPIFKFGGGLTG 1200  
Db 1048 ARRITEVWGIPSPIDTVPTRCNCEGKTPFFKDSQVWRFTNDIKDAGYKPIFKFGGGLTG 1107  
QY 1201 QIVAAALSTAKYNWPSVYFFKRGSGIOQYIYKQEPVOKCPGRPALNYPYVGMTQVRR 1260  
Db 1108 QIVAAALSTAKYNWPSVYFFKRGSGIOQYIYKQEPVOKCPGRPALNYPYVGMTQVRR 1167  
QY 1261 RFRERAIGPSOHTIRIOYSPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPNIRK 1320  
Db 1168 RFRERAIGPSOHTIRIOYSPARLAYQDKGLVHNEVKVSIILWRGLPNVVTSAISLPNIRK 1227  
QY 1321 PDGYDYAFSKQYNYNDVPSRTARAITTRSGQTLKSVWYNCP 1363

Db 1228 PDGYDYAFSKQYNYNDVPSRTARAITTRSGQTLKSVWYNCP 1270  
RESULT 14  
US-07-757-022B-42  
; Sequence 42, Application US/07757022B  
; Patent No. 6431142  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1311 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-42  
Query Match 93.2%; Score 6780.6; DB 4; Length 1311;  
Best Local Similarity 90.5%; Pred. No. 1.2e-182;  
Matches 1270; Conservative 0; Mismatches 0; Indels 134; Gaps 2;  
QY 1 MAWKTLPIYLILLLSVFVIQVSSQ----- 25  
Db 1 MAWKTLPIYLILLLSVFVIQVSSQDLSSCAGRGEGYSRDATCNCYNCQHYMECCPDF 60  
QY 26 -----ELSCKGRCFESFERGECDDCAQCKYDKCCPDYEFCAEVHNPTSPSSKKAP 79  
Db 61 KRVTCTAELSCKGRCFESFERGECDDCAQCKYDKCCPDYEFCAE----- 106  
QY 80 PPGASQTIKSTTKRSPKPPNKKTKVIESEEITEHSVSENQESSSSSSSSSSSTIW 139  
Db 107 ----- 106

QY	140	KI KSSKNSAANRELOKKLVXVONKKNKURTKKQTPPKPVVDEAGSGLNDGDFKVTTPTDST	199
DB	107	-----VKONKKURTKKKTPPKPVVDEAGSGLNDGDFKVTTPTDST	147
QY	200	TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSITVKNKETTIVETKETTITNNKQISTDG	259
DB	148	TQHNKVSTSPKITTAKPINPRPSLPPNSDTSKETSITVKNKETTIVETKETTITNNKQISTDG	207
QY	260	KEKTTSAKETQSI EKTSAKDLAPT SKVLAKPTPKAEITTKGPAITTPKBPRTTTPKBPAS	319
DB	208	KEKTTSAKETQSI EKTSAKDLAPT SKVLAKPTPKAEITTKGPAITTPKBPRTTTPKBPAS	267
QY	320	TTPKBPTTTIKSAPTTPKEPAPTTTKSAPTTPKBPATTTTKEPAPTTTPKBPATTTKEP	379
DB	268	TTPKBPTTTIKSAPTTPKEPAPTTTKSAPTTPKBPATTTTKEPAPTTTPKBPATTTKEP	327
QY	380	APTITTKSAPTTPKBPATTPPKAPTTTPKEPAPTTTPKBPRTTTPKEPAPTTTKEPAPTTPK	439
DB	328	APTITTKSAPTTPKBPATTPPKAPTTTPKEPAPTTTPKBPRTTTPKEPAPTTTKEPAPTTPK	387
QY	440	EPAPTAPKBPATTPKBPATTPKBPATTTTKEPSPPTTPKBPATTTTKSAPTTTKEPAPT	499
DB	388	EPAPTAPKBPATTPKBPATTPKBPATTTTKEPSPPTTPKBPATTTTKSAPTTTKEPAPT	447
QY	500	TTKSAPTTTPKBPSPPTTTKEPAPTTTPKBPATTPPKKAPTTTPKEPAPTTTPKBPATTTTKKP	559
DB	448	TTKSAPTTTPKBPSPPTTTKEPAPTTTPKBPATTPPKKAPTTTPKEPAPTTTPKBPATTTTKKP	507
QY	560	APTAPKBPATTPKETAPTTPKKLTPTTTPKELAPTTTPKBPAPTTPEELAPTTPEBPTPTT	619
DB	508	APTAPKBPATTPKETAPTTPKKLTPTTTPKELAPTTTPKBPAPTTPEELAPTTPEBPTPTT	567
QY	620	PEBPATTPKAAANNTTPKEPAPTTTPKBPATTPKBPATTTTPKEAPTTTPKGTAPTTTLKEP	679
DB	568	PEBPATTPKAAANNTTPKEPAPTTTPKBPATTPKBPATTTTPKEAPTTTPKGTAPTTTLKEP	627
QY	680	APTTPKBPAPKELAPTTTKEPTSTTSDDKAPATTPKGTAPTTTPKBPATTPKBPATTPKPG	739
DB	628	APTTPKBPAPKELAPTTTKEPTSTTSDDKAPATTPKGTAPTTTPKBPATTPKBPATTPKPG	687
QY	740	TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSDKAPATTPKETAPTTTPKEPAPTTPK	799
DB	688	TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSDKAPATTPKETAPTTTPKEPAPTTPK	747
QY	800	KPAPTTEPTBPPTTSEVSTPTTTKEPTTIHKSPDESTEPLSAEBTPPKALENSKPEGVPT	859
DB	748	KPAPTTEPTBPPTTSEVSTPTTTKEPTTIHKSPDESTEPLSAEBTPPKALENSKPEGVPT	807
QY	860	TKTPAAATPKBMTTTAKOKTTERDLRTPPTTAAPKMTKETATTTEKTTTSKIIATTQV	919
DB	808	TKTPAAATPKBMTTTAKOKTTERDLRTPPTTAAPKMTKETATTTEKTTTSKIIATTQV	867
QY	920	TSITTOPTTPKITTTLTKITTLAPKVTTTKKITTTEIMNKPEETA PKDRATNSKATTPK	979
DB	868	TSITTOPTTPKITTTLTKITTLAPKVTTTKKITTTEIMNKPEETA PKDRATNSKATTPK	927
QY	980	PQKPTKAPKAPTSTPKBPTMPRVRKPKTTTPPKMTSTMPELNPTSIASAMLOTTTRPN	1039
DB	928	PQKPTKAPKAPTSTPKBPTMPRVRKPKTTTPPKMTSTMPELNPTSIASAMLOTTTRPN	987
QY	1040	QTPNSKLVEVNPXSDEAGGAETPHMLLRPHVFMPEVTPMDVLPVPMNOGIIINPMLS	1099
DB	988	QTPNSKLVEVNPXSDEAGGAETPHMLLRPHVFMPEVTPMDVLPVPMNOGIIINPMLS	1047
QY	1100	DETNI CNKGKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITTEWGI PSPIDTVFT	1159
DB	1048	DETNI CNKGKPDVGLTTLRNGTLVAFRGHYFWMLSPPSPSPARRITTEWGI PSPIDTVFT	1107
QY	1160	RCNCEGKTFPKOSQYWRFTNDIKDAGYPKPIPKGFGGLTGQIVAAALSTAKYKNWPBSVY	1219
DB	1108	RCNCEGKTFPKOSQYWRFTNDIKDAGYPKPIPKGFGGLTGQIVAAALSTAKYKNWPBSVY	1167

Qy	1220	FFRGGSIOQYIYKQBPVQCPGRRPALNPVYVGMTQVRRRFRERAIGPSQTHTRIQY	1279
Db	1168	FFRGGSIOQYIYKQBPVQCPGRRPALNPVYVGMTQVRRRFRERAIGPSQTHTRIQY	1227
Qy	1280	SPARLAYQDKGVLHNEVKVSIILWRGLPNVUTSAISLNPTRKPDGYDYAFAFKDQYYNIDV	1339
Db	1228	SPARLAYQDKGVLHNEVKVSIILWRGLPNVUTSAISLNPTRKPDGYDYAFAFKDQYYNIDV	1287
Qy	1340	PSRTARAITTRSGOTLSKWYNCP	1363
Db	1288	PSRTARAITTRSGOTLSKWYNCP	1311

RESULT 15

```

US-07-757-022B-104
; Sequence 104, Application US/07757022B
; Patent No. 6433142
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 104:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1140 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-104

```

Query Match 80.0%; Score 5820.9; DB 4; Length 1140;  
Best Local Similarity 96.4%; Pred. No. 9.2e-156;  
Matches 1099; Conservative 0; Mismatches 0; Indels 41;

QY	1	MAWKTLPIYLLLLSVFVIQVSSQ	-----	25
----	---	-------------------------	-------	----

Search completed: October 13, 2004, 11:59:15  
Job time : 29.693 secs

Db	1	MAWKTLPIYLLLLSVFVIQVSSQDLSSCAGCGEGYSRDATA	CNCYNCQHWMCCPDF	60
QY	26	-----ELSCGRCFPSFGRGRCDCDAQCKYDKCCPDYESP	CAVHNPTSPSSKKAP	79
Db	61	KRVCTAELSCGRCFPSFGRGRCDCDAQCKYDKCCPDYESP	CAVHNPTSPSSKKAP	120
QY	80	PPSGASQTIKSTTKRSPKPNKKTKKVIESEIITEHSVSEN	QESSSSSSSSSTIW	139
Db	121	PPSGASQTIKSTTKRSPKPNKKTKKVIESEIITEHSVSEN	QESSSSSSSSSTIW	180
QY	140	KIKSSKNSAANRELQKKLVKDKNKNRTKKPTPKPVVDEAG	SLDNGDFKVTTPDTST	199
Db	181	KIKSSKNSAANRELQKKLVKDKNKNRTKKPTPKPVVDEAG	SLDNGDFKVTTPDTST	240
QY	200	TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETT	VETKEITTTNKQISTDG	259
Db	241	TQHNKVSTSPKITTAKPINRPSLPNSDTSKETSITVNKETT	VETKEITTTNKQISTDG	300
QY	260	KEKTTSAKETQSIKTSKADLAFTSVLAKPTKAETTTKGPAL	TTTTPKEPTTTPKEPAS	319
Db	301	KEKTTSAKETQSIKTSKADLAFTSVLAKPTKAETTTKGPAL	TTTTPKEPTTTPKEPAS	360
QY	320	TTPEKEPTTTIKSAPITTPKEPAPTTTKSAPITTPKEPAP	TTTKBPAPTTTPKEPAPTTTKKEP	379
Db	361	TTPEKEPTTTIKSAPITTPKEPAPTTTKSAPITTPKEPAP	TTTKBPAPTTTPKEPAPTTTKKEP	420
QY	380	APTTTKSAPITTPKEPAPTTPKKAPITTPKBPAPTTTPKE	PTTPKBPAPTTTPKEPAPTTPK	439
Db	421	APTTTKSAPITTPKEPAPTTPKKAPITTPKBPAPTTTPKE	PTTPKBPAPTTTPKEPAPTTPK	480
QY	440	EPAPTAPKKAPITTPKEPAPTTTPKEPAPTTTKBPSPIT	TPKEPAPTTTKSAPITTKBPAPT	499
Db	481	EPAPTAPKKAPITTPKEPAPTTTPKEPAPTTTKBPSPIT	TPKEPAPTTTKSAPITTKBPAPT	540
QY	500	TTKSAPITTPKEPAPTTTPKEPAPTTTPKKAPITTPKBP	APTTTPKBPAPTTTPKEPAPTTTKKP	559
Db	541	TTKSAPITTPKEPAPTTTPKEPAPTTTPKKAPITTPKBP	APTTTPKBPAPTTTPKEPAPTTTKKP	600
QY	560	APTAPKEPAPTTPKETAPTTPKKLTPTTPPKLAPTTPKE	KAPITTPPELAPTTPKEPTPTT	619
Db	601	APTAPKEPAPTTPKETAPTTPKKLTPTTPPKLAPTTPKE	KAPITTPPELAPTTPKEPTPTT	660
QY	620	PEEPAPTTPKAAANTPKBPAPTTTPKEPAPTTTPKETAP	TPPKGTAPTTLKEP	679
Db	661	PEEPAPTTPKAAANTPKBPAPTTTPKEPAPTTTPKETAP	TPPKGTAPTTLKEP	720
QY	680	APTTPKKPAPKELAPTTTKETSTTSKAPAPTTPKGTAPT	TPPKBPAPTTTPKEPAPTTTPKG	739
Db	721	APTTPKKPAPKELAPTTTKETSTTSKAPAPTTPKGTAPT	TPPKBPAPTTTPKEPAPTTTPKG	780
QY	740	TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPAP	TTPKETAPTTPKEPAPTTPK	799
Db	781	TAPTTLKEPAPTTPKKAPKELAPTTTKGPTSTTSKAPAP	TTPKETAPTTPKEPAPTTPK	840
QY	800	KPAPTTPPTTSEVSTPTTTKEPTTIIHKSPDESTPPELS	APTEPKALENSPKBPVGPT	859
Db	841	KPAPTTPPTTSEVSTPTTTKEPTTIIHKSPDESTPPELS	APTEPKALENSPKBPVGPT	900
QY	860	TKTPAAKPEMTTAAKOTTERDLRTPETTTAAAPKMTKET	ATTTTESKITATTQV	919
Db	901	TKTPAAKPEMTTAAKOTTERDLRTPETTTAAAPKMTKET	ATTTTESKITATTQV	960
QY	920	TSTTTQDTPPKIITTLKTTLLAPKVTITTKIITTEINMK	PEETAKPKDRATNSKATTPK	979
Db	961	TSTTTQDTPPKIITTLKTTLLAPKVTITTKIITTEINMK	PEETAKPKDRATNSKATTPK	1020
QY	980	POKTPKAPKPTSTKKPKTMPVRKPTTTPRKNVSTMPPEL	NPTSRIAEAMLOTTTRPN	1039
Db	1021	POKTPKAPKPTSTKKPKTMPVRKPTTTPRKNVSTMPPEL	NPTSRIAEAMLOTTTRPN	1080
QY	1040	QTPNSKLVENPKSEDAGAEGETPHMLLRPHVFMFEVTPD	MDYLPVRVNOGIIINPMLS	1099
Db	1081	QTPNSKLVENPKSEDAGAEGETPHMLLRPHVFMFEVTPD	MDYLPVRVNOGIIINPMLS	1140